DOCUMENT RESUME

ED 214 998 TM 820 261

AUTHOR	King, Jean A.; And Others
TITLE	Improving Evaluation Use in Local School Settings. Optimizing Evaluation Use: Final Report.
INSTITUTION	New Orleans Public Schools, LA. Dept. of Research and Evaluation.
SPONS AGENCY	National Inst. of Education (ED), Washington, D.C.
PUB DATE	Jun 81
GRANT	NIE-G-80-0082
NOTE	98p.; Based partially on a paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, 1981).
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; *Evaluation; Evaluators; Literature Reviews; *Research Utilization; *School Districts
IDENTIFIERS	Evaluation Problems; *Evaluation Utilization

ABSTRACT

A project for studying ways to optimize utilization of evaluation products in public schools is reported. The results indicate that the negative picture of use prevalent in recent literature stems from the unrealistic expectation that local decision-makers will behave in a classically rational manner. Such a view ignores the political settings of real world evaluation use where the non-use of results may constitute a viable and "rational" alternative. Empirical research has suggested that results are used and in a variety of ways. Factors that have not been demonstrated to affect use include the methodological quality of the evaluations and the timeliness of its report. Factors that do seem to affect use include the report source, content, and receiver; political considerations; and the "personal factor," i.e. the presence of at least one person who cares about the evaluation and its results. A collaborative approach to evaluation capitalizes on the factors that have been shown to affect use, although the practicability of such an approach at the local level may make it unfeasible. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED214998

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

OPTIMIZING EVALUATION USE: FINAL REPORT

A National Institute of Education Teaching and Learning Grant

NIE-G-80-0082

IMPROVING EVALUATION USE IN LOCAL SCHOOL SETTINGS

NEW ORLEANS PUBLIC SCHOOLS
Department of Research and Evaluation
June, 1981

TM 820 261

IMPROVING EVALUATION USE IN LOCAL SCHOOL SETTINGS

Jean A. King
Tulane University

Bruce Thompson
University of New Orleans

Ellen M. Pechman
New Orleans Public Schools

June, 1981

FINAL REPORT: OPTIMIZING EVALUATION USE
NIE Teaching and Learning Grant
NIE-G-80-0082

Based in part on a paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, 1981. The development of this paper was partially supported by a grant from the National Institute of Education; however, no endorsement by the Institute is implied.

Foreword

This paper is the final report of a 1980-1981 NIE grant to the Orleans Parish School Board (NIE-G-80-0082) for studying ways to optimize the utilization of evaluation products in public schools. Most of the literature discussed here was compiled from July to November, 1980. Our compilation was aided greatly by two earlier bibliographies prepared at Northwest Regional Educational Laboratory (Smith, 1980a; Hansen and Martin, 1980). In addition, we conducted an ERIC computer search and reviewed the Education Index evaluation listings from 1975 on and the American Educational Research Association's Annual Meeting program and abstracts for 1978, 1979, and 1980. The resulting bibliographies are available through ERIC in two forms--an extensive bibliography of 326 entries (King, Thompson, & Pechman, 1981a) and a shorter bibliography annotating roughly 20 important pieces (King, Thompson, & Pechman, 1981b). To be included, papers, articles, chapters, and books had to meet two criteria:

1. They either had to address directly the topic of the utilization of program evaluation information at the local education agency (LEA) level and/or have been extensively cited in the literature and hence be important conceptually. This eliminated much of the body of literature on policy analysis and evaluation use at the federal level, as well as literature related to other evaluation settings (e.g., public health, social work, hospitals). This also eliminated work concerning the use of research results or of knowledge in general and on the

evaluation of teaching performance.

2. The literature had to be relatively current and readily available. Anything ten years or older was included only when it was seminal or had clearly influenced later thinking in the field. Pieces that we were unable to locate after extensive checking were eliminated.

Any bibliography in a rapidly changing field like evaluation must be regarded as dated even before it appears. For this reason, some of the literature discussed in the following paper is not cited in the earlier bibliographies, and readers are additionally encouraged to use the bibliography at the end of this paper in their examination of the evaluation use literature.

For the purpose of this paper, the term evaluation has been broadly defined. The question of definition is an important one since what one counts as evaluation information is clearly affected by one's definition of evaluation. Webster and Stufflebeam (1978; see also Stufflebeam and Webster, 1980) identify 13 alternative approaches to evaluation, grouped in the three categories of pseudo-evaluation, quasi-evaluation, and true evaluation. Given these categories, the evaluation process can involve radically different activities, and "evaluation" results can range from a carefully manipulated public relations campaign to an empirical generalization published in an educational journal; from a national policy to a detailed case study; from a staff brain-storming session to the administration of a test.

In addition to this categorization are numerous others in the field, for example, the distinctions between formative and

summative evaluation (Scriven, 1967; Alkin et al., 1974) and among the research, decision-making, and values assessment components of evaluation (Smith, 1980c). It is beyond the scope of this paper to discuss the implications of the various approaches and components of evaluation. For this reason, the following definition of evaluation is stipulated: "the process of providing information to administrators to help them make decisions regarding educational programs." This definition is purposely broad, encompassing pseudo-, quasi-, and true evaluation. Such a definition is needed to discuss evaluation use in local settings where intuitive definitions abound, often equating evaluation with achievement testing (David, 1981). To limit the definition to Webster and Stufflebeam's notion of true evaluation is to lose sight of many examples of use. The decision-making focus of the definition relates to this project's purpose, i.e. to determine strategies for increasing evaluation use by LEA personnel. Following a conceptual discussion, the paper is divided into a summary of empirical work and a discussion of other writings on ways to increase the use of evaluation information.

Several names can apply to local users of evaluation information. Throughout the paper, the terms decision-maker, administrator, and information user are used interchangeably. Also, following Weiss (1979), the term use, rather than utilization, will be used. As Weiss writes, utilization "embodies an inappropriate imagery . . . because of its overtones of instrumental episodic application. People don't utilize

research the way that they utilize a hammer" (p. 2).

The authors wish to thank Robert Brown, James Sanders, and Michael Vitale, who wrote critiques of a draft of this paper, for their valuable and insightful comments.

TABLE OF CONTENTS

Foreword. . . . . page iii

Abstract. . . . . page ix

Executive Summary . . . . . page x

Introduction. . . . . page 1

The Rationality of Use and Non-use. . . . . page 4

The Concept of Use. . . . . page 12

Empirical Studies of Local Evaluation Use . . . . . page 26

Strategies for Improving Use. . . . . page 50

A Critique of Previous Use Research . . . . . page 67

Research Agenda for the Future. . . . . page 71

Making Local Evaluation Use Happen. . . . . page 73

Figures and Tables

Figure 1. The relation of evaluation use to evaluation impact . . . . . page 19

Figure 2. Types of use and related actions . . . . . page 25

Table 1. Issues raised by users in Kennedy, Apling, and Neumann (1980) . . . . . page 35

OPTIMIZING EVALUATION USE

(NIE-G-80-0082)

Abstract

The purpose of the project was to facilitate optimal use of evaluation research by generating knowledge regarding those factors that affect utilization.

The project has resulted in three products: (1) a compilation of previous theorizing and research related to local evaluation use, including an extensive bibliography, an annotated bibliography, and an integrated literature review; (2) a survey of local administrators' perceptions of evaluations; and (3) a report simulation study of local users' perceptions of evaluators.

The results indicate that the negative picture of use prevalent in recent literature stems from the unrealistic expectation that local decision-makers will behave in a classically rational manner. Such a view ignores the political settings of real world evaluation use where the non-use of results may constitute a viable and "rational" alternative. Empirical research has suggested that results are used and in a variety of ways. Factors that have not been demonstrated to affect use include the methodological quality of the evaluations and the timeliness of its report. Factors that do seem to affect use include the report source, content, and receiver; political considerations; and the "personal factor," i.e. the presence of at least one person who cares about the evaluation and its results. A collaborative approach to evaluation capitalizes on the factors that have been shown to affect use, although the practicability of such an approach at the local level may make it unfeasible.

OPTIMIZING EVALUATION USE

(NIE-G-80-0082)

Executive Summary

For the past ten years the literature on evaluation use has lamented the fact that, despite millions of dollars being spent on educational evaluation, the results were not being used by local decision-makers. Such non-use is costly because it wastes effort and money, but more importantly because it may mean children are being served by less than optimally effective programs. Our project sought to do three things: (1) to compile and integrate the extant literature on local evaluation use; (2) to survey administrators nationwide as to their feelings about evaluation; and (3) to conduct a report simulation study to see if local administrators perceived differences among evaluators.

The project's first product is a state-of-the-art paper entitled "Local Evaluation Use: A Literature Review and Research Agenda" (King and Thompson, 1981b). The paper makes several major points. First, decision-making in the real world of public schools is not a rational process in the traditional sense. Local users of evaluation information engage in "satisficing," i.e. in finding solutions that are "good enough" given constantly changing and uncertain circumstances. Second, and related to this, is the fact that the non-use of evaluation results must not be viewed as irrational or non-rational; in an important sense, non-use represents a viable form of use in the political contexts of local evaluation settings. Third, then, the concept of use must be broadened from the limited notion of instrumental use (i.e. where a result suggests a specific action and that action is taken) to include a variety of activities. A necessary condition for use is the conscious employment of results by a decision-maker for whatever purpose; without such awareness and choice on someone's part, the activity that ensues is not use, but rather, to use Smith's (1980) term, evaluation impact. Theorists have generated several frameworks for thinking about evaluation use, supporting the broadened definition.

Fourth, few empirical studies of local evaluation use can be found in the literature. Most writing that has been done on the topic is informal and anecdotal. The research that has been done has centered around three questions: (1) Are evaluation results used? (2) What are the characteristics of local evaluation units and users? and (3) What factors affect use at the local level? The answer to the first question is definitely yes, if one adopts the broadened definition of use discussed above. Local administrators <u>do</u> use evaluation results, although not necessarily in predictable ways.

The answers to the second and third questions are far more complex. Existing research has only begun to provide insight into the process of use at the local level. Regarding the second question, two generalizations can be made: first, local evaluation units are typically not involved in program improvement; and second, although some research indicates that local users find evaluation helpful, other data suggest that these users may feel that evaluation is not of value to them. The picture of the local evaluation context currently available suggests reason for the non-use of evaluation information.

The factors affecting use at a local level can be divided into those

that evaluators control and those that are part of the evaluation context and must therefore be worked with. Research suggests that methodological quality does not necessarily lead to the use of results, although good evaluation practice nevertheless stresses its importance. Variations in evaluation reports have been shown to affect use; the message source, content, and the characteristics of the receiver can all affect what happens to results. Interestingly, the variable of timeliness, often cited as a critical factor for use, has not been shown to be significant, probably because of the informal contacts between evaluators and clients that make the evaluation report a mere formality.

Research has also pointed to the importance of certain environmental and contextual factors affecting use. The issues faced by various users differ significantly, and the value of evaluation information for addressing some problems (e.g., rapidly changing issues at the district-level) is, to date, unclear (Kennedy, Apling, and Neumann, 1980). Two empirically validated factors are the "political considerations factor" and the "personal factor," i.e. the presence of at least one person in an evaluation who cares about the process and its results (Patton, 1978). Given these issues and factors, the research literature repeatedly recommends a collaborative approach to evaluation, despite the fact that such an approach may be extremely time-consuming and has not been empirically demonstrated to be effective.

The next points the paper makes are based on the anecdotal literature that by and large has found validation in empirical work. Suggestions for improving evaluation reports include using multi-media presentations, educating the press, writing executive summaries, including carefully worded policy recommendations and so on. The importance of informal contacts between evaluators and clients stressed in anecdotal literature suggests that the formal report itself may be less important. Again, the collaborative approach finds support despite inherent problems like the "goals shuffle," whereby local staff decide to switch goals half-way through an evaluation, or the "methodology shuffle," whereby users rent statisticians to attack any results that are unpopular.

The non-empirical evaluation literature suggests numerous approaches for successful collaboration. Recommendations include the following: targeting evaluations, identifying issues of interest to users, employing informal reporting procedures, anticipating users' future questions, and showing users that the evaluator cares. A further approach of particular merit for local users is the process called evaluability assessment (see, for example, Rutman, 1980) which determines when and if a program is ready to be evaluated.

The final point made in the paper is an obvious one: we have much to learn about the use of evaluation information at the local level. Previous research has been dominated by retrospective case studies and simulation studies, both of which have limitations as well as strengths. Research is needed in several directions, both conceptual and practical. Most needed, perhaps, is prospective research on the process of evaluation use, i.e. current case studies, and research on the outcomes of effectiveness training for evaluators.

Because they were empirical studies, the second and third project products can be summarized more briefly. The nationwide survey of local administrators (King and Thompson, 1981a) found first, that the majority of users surveyed--60%--reported that the evaluations of educational programs in their schools were either useful or very useful. Second, regardless of district size, most

users reported that they do use the information they receive in their decision-making. The survey also revealed two less heartening results: well over two-thirds of those reporting (72%) felt that the program effects they most cared about could only be measured indirectly (42%) or not at all (30%); and few local users have much contact with evaluators. Given these differing perspectives on measurement and limited contact between evaluators and users, the positive attitude and use of results mentioned earlier may be indicative of evaluation's power.

The third product of the project was a study entitled "Evaluation Types: Do Differences Matter?" (Thompson and King, 1981). The study investigated whether administrators implicitly perceive the four types of evaluators proposed by Meltsner (1976) (i.e. entrepreneur, politician, technician, and placeholder) when they consider evaluative information. One simulated evaluation report was written to represent each evaluator type, then two pilot studies were conducted to assess the validity of the simulation materials. The results suggest that administrators may implicitly employ the typology; they were particularly attentive to the technical merit of reports.

REFERENCES

Kennedy, M., Apling, R., & Neumann, W.F. The role of evaluation and test information in public schools. Cambridge, MA: The Huron Institute, 1980.

King, J.A., & Thompson, B. A nationwide survey of administrators' perceptions of evaluation. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1981.

King, J.A., Thompson, B., & Pechman, E.M. Improving evaluation use in local school settings. Final report: Optimizing evaluation use (NIE-G-80-0082). New Orleans Public Schools, New Orleans, 1981.

Meltsner, A.J. Policy analysts in the bureaucracy. Berkeley: University of California Press, 1976.

Patton, M.Q. Utilization-focused evaluation. Beverly Hills: Sage, 1979.

Rutman, L. Planning useful evaluations. Beverly Hills: Sage, 1980.

Smith, N.L. The president's corner. Evaluation news, Summer, 1980, 24-25.

Thompson, B., & King, J.A. Evaluator types: Do differences matter? Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, 1981.

Introduction

Reading the literature on the use of program evaluation information may discourage even stout-hearted evaluators. A sampling of quotations from the last ten years is sufficient to document the widespread feeling among evaluation commentators that program evaluation results in general are not having the effect they were intended to have. Wholey et al. (1970) concluded that "the recent literature is unanimous in announcing the general failure of evaluation to affect decision-making in a significant way" (p. 46). In a 1972 paper on evaluation utilization, Weiss wrote that "evaluators complain about many things, but their most common complaint is that their findings are ignored" (p. 319). Writing a year later, Worthen and Sanders (1973) noted that "evaluation is one of the most widely discussed but little used processes in today's educational systems" (p. 1). The same year, Rippey (1973) concluded that

> At the moment, there seems to be no evidence that evaluation, although the law of the land, contributes anything to educational practice other than headaches for the researcher, threats for the innovators, and depressing articles for journals devoted to evaluation (p. 9).

Five years later Patton (1978) wrote that

> In many ways the odds are all against utilization and it is quite possible to become skeptical about the futility of trying to have impact in a world where situation after situation seems impervious to change (p. 291).

Recently, Alkin and Daillak (1979) stated that "there have been great hopes for evaluation, not only among evaluators themselves, but also among other educators, elected officials, and the public.

Yet these hopes have dimmed" (p. 41). The uneasy feeling that evaluation results are not being used has not changed. As Stake (1976) once wrote, "We do not know whether or not evaluation is going to contribute more to the problems of education or more to the solutions" (p. 1).

This characterization apparently applies equally well to judicial (Saks, 1980) and legislative settings (Brandl, 1980; Mitchell, 1980), and, unfortunately, to local education agencies as well. Within the last two years, three writers involved in public school settings have written the following:

> In an ideal world we wouldn't have to worry about utilization. Educators would be eagerly awaiting our findings and would promptly rush to put them into practice. I don't need to tell you that isn't happening (Holley, 1979, p. 2).

> . . . although the potentiality exists, local use of evaluation . . . does not occur routinely as a natural consequence of conducting an evaluation (Kennedy, Apling, & Neumann, 1980, p. 5).

> All LEAs, with possibly a few exceptions, can point to their volumes of research and evaluation verbiage sitting on the shelves of district administrators being used for little else than a door stop, swatting flies, or any of the other various and sundry purposes for which research is used in the public schools (Kilbourne & DeGracie, 1979, p. 12).

The consensus in the literature is virtually unanimous: the results of program evaluations are having little effect in the very places they were meant to.

Although these commentators are expressing largely undocumented personal feelings, the concern is a valid one. Alkin et al. (1974) have written that "if evaluative information is not useful, if it does not serve the needs of decision makers, then

evaluation has lost its justification" (p. 1). To the extent that evaluative results that could have a positive effect go unused, the entire evaluation enterprise in some sense fails.

The non-use of evaluation findings when that use would be appropriate is costly in three ways. First, non-use represents an enormous waste of effort. As Datta (1979) notes,

> Considerable effort is involved in conducting almost any evaluation: in identifying the evaluation question, in designing the study, in overcoming the obstacles to conducting an evaluation and protecting it methodologically from uninterpretability (p. 22).

Second, non-use represents a waste of monies. At the federal level, for example, direct expenditures on non-defense evaluation projects in 1974 by the federal government alone amounted to $146 million (Kelezo, 1974). Surprisingly, at the very time theorists and users alike are questioning evaluations' results and effects, money continues to be invested in evaluation studies. The indirect costs of non-use, when ineffective programs are not modified or discontinued, are even more staggering at both the local and the federal levels.

The third and most serious cost of non-use, however, is the human cost: non-use means that the clients of educational programs receive less than optimally effective service. Failure to use evaluative information is especially unfortunate because, as Wise (1980) notes, "No one else is given the resources and time to question, observe, assess, weigh, probe, and reflect that the evaluator is given" (p. 16). At the local level, the costs of non-use can weigh heavily.

This paper is a review and integration of the literature on

evaluation utilization, with special reference to the use of evaluation findings at the local level. Following a discussion of the rationality of use and non-use, the paper will present a conceptual framework for evaluation use, a review of empirical research on the topic, and a discussion of observations and recommendations from the literature. The final section of the paper will include a critique of previous research and a research agenda for further investigation of local evaluation use.

The Rationality of Use and Non-Use

Well before the passage of the Elementary and Secondary Education Act in 1965 signalled a new era for program evaluation, educators paid homage to a myth concerning the presumed place of evaluation in the world of education. The myth went something like this:

> Harried decision-makers, over-worked but rational at heart, are presented evaluative information that is relevant to decisions they are about to make. Sitting calmly at their desks, they consider the data, weigh their options, then make the "correct" decision, i.e. the choice supported by the data. The evaluator, beaming happily behind the scenes, takes pride in nurturing such rational action.

The series of quotations given at the beginning of this paper suggests that this myth--what Andrews (1979) calls the "big bang" theory of evaluation--has rightfully been exploded. First, such a view denies the process and political context of real world decision-making, and second, it ignores the rationality of non-use.

Decision-making in the Real World. Perceptions of how decisions are made clearly affect judgments of the extent and

quality of the use of evaluative information. As Wise (1978) explains,

> Referring to administrators as 'decision-makers' and to what they do as 'decision-making' may have been a first step in creating the utilization problem, for we expect to see decisions being made by someone called a decision-maker (p. 6).

Thus, many evaluators presume that evaluation ought to be used merely because its evidence is rational. As Cronbach et al. (1980) note, however, "If the term decision is understood to mean a formal choice at a particular time between discrete alternatives, decision making is rarely to be observed" (p. 84).

Patton (1978) reviews three competing conceptions of organizations that clarify the process of decision-making in the real world: the rational goal attainment model; the systems model; and the open systems perspective (pp. 122-127). Because the open systems perspective most fits the context of educational decision-making, decision-makers in these settings can be viewed as engaged in "satisficing," i.e., the process "of finding a course of action that is 'good enough'" (Simon, 1957, p. 204). As Patton (1978) points out, the decision-maker in such a setting can never have all the information needed to meet the conditions for rational action and even if he or she could, it would be beyond an individual's capability to comprehend it all (p. 125). "Satisficing" is then tenable because administrators feel they can change most courses of action if decisions later prove to be seriously mistaken.

Some administrators perceive evaluation as a two-edged sword. The pretense of objectivity embues evaluation with

credibility. The administrator who is confronted with an unfavorable result knows that any result can be attacked on several grounds. On the other hand, the administrator who is confronted with a favorable result may perceive the result as a weapon against adversaries. Firm stances, however, may not be taken until the results dictate them. From the perspective of satisficing, evaluative information can thus be useful in a variety of ways not necessarily predictable from a "rational" view.

Patton (1978), summarizing the results of an empirical use study, writes,

> The kind of impact we found, then, was that evaluation research provided some additional information that was judged and used in the context of other available information to help reduce the unknowns in the making of difficult decisions (p. 30).

Cronbach et al. (1980) distinguish between the context of command, i.e. "of concentrated power and responsibility," and the context of pluralistic accommodation (p. 83). These contexts are useful in describing the likely effects of evaluation results.

> In a context of accommodation, the evaluator cannot expect a "go/no go" decision to turn on his assessment of outcomes, whereas information on outcomes is influential in a context of command--a decision all but made can still be unmade (p. 151).

To ignore the process and context of decision-making in the real world is to be unable to adequately examine the use of evaluation information.

"Rational" non-use. More than likely, as Caplan (1980) writes, "There is real danger in uncritically accepting utilization as desirable or in being oversold on its value. Not

all utilization is good and not all nonutilization is bad" (p. 5). To label a decision-maker "irrational" or "non-rational" because, for whatever reason, he or she does not use the results of a pertinent evaluation is simply inappropriate. In other words, the non-use as well as the use of evaluative information can be highly "rational" in a specific setting and for a number of reasons.

One reason practitioners may "rationally" feel that evaluation studies do not warrant use is that decision-makers may see the world from a different perspective than the evaluator, a perspective that may imply a different form of rationality. Caplan's "Two Communities" theory that addresses the contrasting perspectives of researchers and policy makers (see, for example, Caplan et al., 1975) suggests that an evaluator's concern for technically correct procedures may conflict with a program staff's implementation plan or with a local decision-maker's information needs. While the evaluator sees a local program as a piece of a educational experiment, an administrator may instead see it as providing essential services to students. Wilensky and Lebeaux (1958) observe that "what the social scientist thinks of as 'objective investigation' the practitioner often takes as 'hostile attack'" (p. 20). Such differences in the evaluator's and the practitioner's perspectives may reduce trust and impede effective communication. As Deal and Rallis (1980) explain,

> Theoretically, collaborative relationships require a shared perspective, high trust and power parity. The shared perspective between knowledge producer and user, however, is often characterized by different perspectives, low trust and an asymmetrical distribution of power (p. 216).

Empirical research makes clear how this situation might occur in public school settings where

> Not only have 42% of them (LEA evaluation unit heads) not taught, but 70% have not run a school. This means that even when evaluation heads have teaching backgrounds, they do not take the typical advancement route to the central office (Lyon et al., 1978, p. 66).

Practitioners may also feel that evaluation results contradict their intuitions regarding program impacts. As Guba (1969) notes, "For decades the evidence produced by the application of conventional evaluation procedures has contradicted the experiential evidence of the practitioner. Innovations have persisted in education not because of the supporting evidence of evaluation but despite it" (p. 1). In David's study of the local uses of Title-I evaluations (1981), for example, practitioners expressed strong feelings that, regardless of what the data showed, their program had succeeded (p. 38). Such situations frequently occur when summative evaluations find "no significant differences" associated with a program in the face of the practitioner's direct experience that the program really did substantially alter classroom life. Shapiro (1973) argues, "While it is important to try to explain negative (i.e. statistically non-significant summative) test results, it is far more important to account for the disparity between the negative test findings and the clear differences observed in classroom behavior" (p. 527). The Joint Committee's Standards A3, "Information Scope and Selection" (pp. 27-31), and D10, "Justified Conclusions" (pp. 135-137), point to the

importance of such accounting in good evaluation practice.

Another reason for the rational non-use of evaluation findings relates to the environment of local school settings. Non-use constitutes a rational choice when institutional, political, or the administrator's own goals that are quite distinct from formal program goals suggest it as a viable course of action. For example, Granville (1977) explains that "a decision maker, in addition to considering whether or not a program has fulfilled its manifest objectives, must also consider its fulfillment of latent objectives, such as enhancing the agency's prestige or expanding its resources" (p. 2). This, theoretically, may not be an ideal basis for decision-making, but in practice administrators clearly have their own survival needs and personal agendas that can affect their use of evaluative information. The desire for personal prestige, power--or merely for keeping one's job--may make non-use imperative.

A related reason why results may rationally go unused is that some evaluation studies were never meant to be used. As Alkin (1976) explains, this is frequently the case when an externally mandated evaluation is involved:

> Many practicing school administrators . . . believe that evaluation is simply an event that leads to compliance with various agency requirements. There is no real expectation that major basic decisions will be made. The name of the game is simply not to get "dinged" by the governmental agency (p. 16).

In a recent article, Alkin (1980a) made the same point using the analogy of a garden party.

> . . . suppose the host . . . should insist that each of the guests periodically rate the quality of the party, or the drinks or the food, etc.--it can't

> really be expected to have much impact. This somewhat peculiar, externally imposed requirement will be tolerated as part of the "price of admission," so to speak, but it won't really change the behavior of individuals (p. 3).

Indeed, deliberate non-use is thoroughly rational if an evaluability assessment suggests that a program is not ready for evaluation (see, for example, Rutman, 1980), but, because it has been mandated, an evaluation must be conducted anyway.

A final reason for non-use is that some evaluation studies are poorly done or technically weak and do not merit use. Guba and Stufflebeam (1970) observe:

> Many researchers make wrong assumptions about what an evaluation study should accomplish, and . . . (then) based on these erroneous assumptions, researchers foist bad advice upon unsuspecting and unsophisticated practitioners. As a consequence, evaluations are usually useless, and practitioners are largely justified in the jaundiced view they typically have taken about evaluation and its utility (p. 6).

Ironically, empirical evidence exists (Alkin et al., 1974, p. 48) linking poor quality evaluations with poor quality programs, the very projects that, theoretically, good evaluations should be able to help most. As the Joint Committee's Standards (1981) would suggest, methodological quality is a minimal requirement for good evaluation practice. Without this, the use of results may be inappropriate.

Evaluators are aware that poor technical quality can occur in many forms. An all too common situation that justifies non-use occurs when evaluators do not attend to a program's actual degree of implementation. As Guttentag and Struening (1975) observe, "Obvious though it may seem, evaluations continue

without either raising or answering the primary question: 'Does the program (even) exist?'" (p. 4). This situation continues despite the availability of several strategies for measuring implementation prior to comparison of results for program participants and non-participants (Revicki & Rubin, 1980), including notably the model developed by Hall and Loucks (1977). Standard D1, "Object Identification," encourages evaluators to examine the object of the evaluation sufficiently so that its "form(s) . . . can be clearly identified" (Joint Committee, 1981, p. 99).

Taken together, then, these reasons for the non-use of evaluation results suggest first that "rationality" for a local decision-maker may differ from more traditional definitions of rationality. Evaluators must recognize that administrators at times may not behave rationally in the classical sense. Administrators' decisions may be rational only when viewed from within their political frame of reference. To assume that administrators will apply the results of evaluation in predictable ways is then unrealistic. Local users may see evaluation as external and irrelevant to the program; the immediacy of other concerns in the local context, the potential mistrust of evaluators, and the difficulty of accepting criticism graciously may cause local decision-makers to use or not use evaluation information in their satisficing. Also, as Weiss (1979) explains, because evaluations can end in unclear, incomplete, or conflicting results, the "rational" choice for a decision-maker is frequently not evident, and to label an

administrator non-rational in such cases makes little sense (pp. 2-7).

These reasons also suggest that, in an important sense, non-use represents a viable form of "use." The major task for evaluators is to maximize the appropriate use of evaluation results. If the non-use of results is appropriate, then non-use is highly rational and should not be condemned. The implicit assumption that use is good and non-use bad (Weiss, 1979, p. 7) must, like the myth of rational evaluation use, be recognized as untrue. Weiss (1980, cited in Kennedy, Apling, and Neumann, 1980) identifies five characteristics of rational decisions: (1) purposiveness; (2) boundedness of time, actors, and events; (3) calculation, i.e., "consideration of the aptness of various actions for dealing with the problem situation and achieving results"; (4) perceived significance; and (5) realization that a decision is being made (pp. 132-133). Given these characteristics, the non-use of evaluation results in educational settings may frequently represent a rational decision.

The Concept of Use

Anyone familiar with local evaluation settings knows that the uses to which evaluation results can be put are as varied as the types of information that can be labeled "evaluation." As mentioned above, in some cases evaluations are conducted only in response to federal mandate, and the non-use of results occurs because the evaluation is an externalized event standing apart from the immediate concerns of program staff. In others, a

decision-maker eager to impress someone may want to be the first among his or her peers to commission an evaluation, and the results are almost beside the point. In still others, a series of evaluation results over a period of time may alter or confirm a user's ideas about a specific program or about a general approach to a problem. The list goes on. Any discussion of the use of evaluation information, then, must first attempt to define the key term use so that instances of use can be separated from those of non-use. Only in this way can we hope to conduct research to determine factors affecting use in order to enhance the effect of evaluation results. As Weiss (1979) has written, "Until we resolve questions about the definition of use, we face a future of non-comparable studies of use and scant hope of cumulative understanding of how evaluation and decision-making intersect" (p. 13, emphasis in original).

The concept of evaluation use has changed dramatically over the past eight years. There is what Alkin et al. (1979) call the "mainstream" viewpoint, that

> . . . contends that evaluations seldom influence program decision-makers and holds out little hope that evaluation will ever break through the barriers to real impact on programs (p. 17).

Implicit in the gloomy portrait of evaluation's failure is an unrealistic, rationalistic point of view that assumes the results of evaluations should create immediate and observable effects. The "alternative" viewpoint--that "evaluations do already influence programs in important and useful ways" (p. 17)--has been induced from the combined results of empirical use studies.

Taken together, these results suggest that a possible reason for the generally negative impression of evaluation use comes from an inappropriate definition of what counts as use. Although the focus of these studies is generally on federally funded programs, there are implications for the discussion of evaluation use at the local level in general. Consider the following studies.

--Alkin, Kosecoff, Fitz-Gibbon, & Seligman (1974) examined 42 ESEA Title VII evaluation studies conducted in 17 states during 1970-1971, measuring impact at both the federal and local levels. Almost every project director--39 of the 42--reported that the formative studies were useful to them (p. 26), and "no project director indicated that he would prefer not to have an evaluator on the project" (p. 28, emphasis in original).

--Patton et al. (1977) interviewed decision-makers and evaluators from 20 federal health evaluation studies. Their results indicated that evaluation findings were used by decision-makers, "but not in the clear-cut and organization-shaking ways that social scientists sometimes believe research should be used" (p. 144). Instead, evaluation results provided another piece of a complicated information puzzle that helped to reduce the uncertainty within which federal decision-makers operated.

--Alkin, Daillak, & White (1979) did intensive case studies of five ESEA Title I or Title IVC programs, locating only one instance of mainstream use, i.e. a case where a formal comparative evaluation was conducted, "an evaluation report was written, the findings were negative, and the program was dropped"

(p. 224). What they did find in each case, however, was an indirect kind of use. Decision-makers stated that the evaluation results influenced them in gradual or general ways.

--Dickey (1979b), studying 47 ESEA Title IVC projects in Minnesota, found a high utilization level; 72% of the decision-makers reported the evaluation results to be "very" or "quite" useful.

--Kennedy, Apling, and Neumann (1980), focusing on local uses, discuss the results of 345 interviews with LEA personnel in 18 districts known for successfully using evaluation and testing information. They describe a variety of use issues in four clusters of applications: district-wide, program-level, building-specific, and clinical.

Brown and Braskamp (1980) summarize the current view of evaluation utilization: ". . . effective utilization does not necessarily mean that any of the recommendations are implemented or that there are any immediately apparent decisions based on the information" (p. 91). This suggests both that the discouraging picture of evaluation use created by the quotations above may stem from a mistaken expectation regarding the nature of use and that the extent of evaluation use may therefore be underestimated in the literature. Wise (1978) has suggested that

> If there is an evaluation utilization problem, it is not that decision-makers do not use the information they receive, it is that evaluators cannot easily see their information being used in the incrementalism of real-world decision-making" (p. 24).

Similarly, Alkin, Daillak, & White (1979) argue that

> Taken together, the studies and our observations and experiences suggest to us that evaluation <u>can</u> make a

> difference, that it does so more often than the published critiques suggest, that some school districts characteristically produce a high proportion of useful evaluations, and that some evaluators have acquired skills that allow them to carry out technically competent and programmatically influential evaluations (p. 16, emphasis in original).

Those instances where evaluation results are directly translated into a decision are sufficient, but not necessary cases of use.

Based on their empirical work, Alkin, Daillak, and White (1979) define an "instance of the utilization of local school program evaluation" as

> Evaluation information considered by a local client, sanctioned local users, or external users as a dominant influence, one of multiple influences, or one of multiple, cumulative influences in making decisions, substantiating previous decisions or actions, or establishing attitudes related to establishment, external funding, local district funding, continuance of a component, curriculum/instructional methods, administrative/personnel operations, or community acceptance of the local school program (p. 232).

This definition reflects the "alternative" notion of evaluation use discussed above, recognizing that the evaluation information may be only one source of information used in a complex setting.

Considerably more direct, Smith's (1980c) definition of evaluation use--the "conscious employment of an evaluation (its processes, products, or results) to achieve some desired end or impact" (p. 25)--reflects the purposeful nature of use implicit in Alkin et al.'s definition. By including the process, product, and results of evaluation in his definition, Smith further suggests the need to distinguish the effects of the evaluation process from the effects of its outcomes. As Cronbach et al. (1980) note,

> . . . whether an evaluation is launched to promote a cause or to report neutrally on events, the measurement procedures and reports can easily have a wholly unanticipated influence on what happens next (p. 27).

Knowing that the process itself can result in changes (e.g., people, knowing they are being evaluated, working hard to look good or perhaps to discredit the evaluation), to study use we must trace the evaluation's effects from its inception until well after its completion. This definition appropriately recognizes the dynamic nature of use.

Smith also distinguishes use from impact, which he defines as "any discernable actions, events, or changes in conditions that are directly influenced by the evaluation, its processes, products, or findings" (p. 25). Given this definition, an evaluation needn't be used to have impact. If, for example, teachers work to oust the superintendent who introduced an accountability system, the evaluation has had an impact, regardless of its outcomes or the use to which the superintendent puts it. This distinction is helpful in examining the varied effects any evaluation can have.

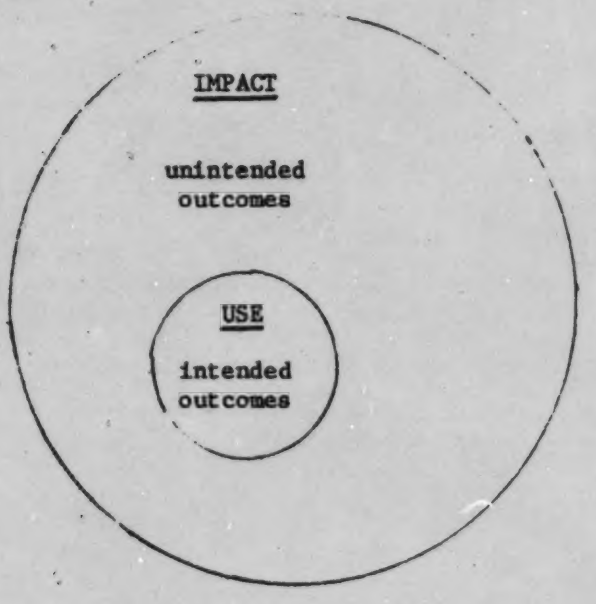
Leviton and Hughes (1979) present two necessary conditions for use: (1) Cook and Pollard's (1977) suggestion that there be "'serious discussion of the results in debates about a particular policy or program'" (p. 4); and (2) ". . . evidence that those engaged in policy or program activities would have thought or acted differently in the absence of the research information" (p. 5). If the notion of "serious discussion" is changed to "serious consideration," following Alkin, Daillak, and White (1979), the

first condition is both a necessary and sufficient condition of evaluation use. When a decision-maker actively and consciously considers how he or she will use an evaluation--even if this is simply an awareness of how an evaluation may frighten certain people--the example is one of use. Eliminate this conscious element, and the example becomes one of impact.

In contrast, the second condition can be considered neither necessary nor sufficient for use at the local level. It was suggested above that non-use may in some cases be an important form of use. Decision-makers may not think or act differently as a result of seriously considering the results of an evaluation; they may react to the findings and choose to ignore them. To the extent that information once received changes a person's thoughts forever, the second condition may be technically "necessary," but because observers may be hard pressed to find evidence of this, it seems pointless to make such a case. Further, the second condition alone is not sufficient to determine an example of use because, as stated above, without the conscious consideration marked by the first condition, the example is one of impact rather than use. Merely seeing activities reflecting change is not sufficient evidence of evaluation use.

Figure 1 presents these ideas in diagram form. A decision-maker's or information user's intentions will result in the use of evaluation in a variety of ways. These are simultaneously examples of evaluation use and impact. All other outcomes, i.e. those unintended by the user, are examples of evaluation impact. Both categories represent impact in that the

Figure 1. The relation of evaluation use to evaluation impact.



evaluation directly influences "actions, conditions, or changes in conditions." Given this conceptualization, the notion of indirect influences is no longer helpful; either a user consciously uses results (an example of use) or, without a user's intention, results create a direct effect (an example of impact).

Consider, for example, an evaluation that indicates serious weaknesses in an ESL program for Hispanic children. Once the results are available in any form, use could include a central administrator's decision to remove the project director, a group of parents' complaint to the school board that their children are being poorly served, or a teacher's decision to fight for the program due to the evaluation's inadequate measurement procedures. For each user, the evaluation may have additional, unintended impacts. The central administrator, for example, having fired the project director, may be dismayed to learn of the parents' complaint because the director may have been an effective community liaison. From this example, it is clear that one person's use can be another's impact.

Several writers have outlined conceptual dimensions of evaluation use (Alkin, 1975; Patton et al., 1977; Alkin, Daillak, and White, 1979; Conner, 1979; Weiss, 1979; Brown, 1981). Weiss (1979) has suggested six dimensions for deciding what constitutes a use that would provide conceptual clarity for studies:

1. What is used?
2. How direct is derivation from the study?
3. By whom is it used?
4. By how many people?
5. How immediate is the use?
6. How much effect is required? (pp. 11-12).

Brown's (1981) conceptualization includes five dimensions: (1) Who uses the information? (2) For what purpose is the information used? (3) What is the evaluation context? (4) To what extent is the information used? and (5) What kinds of information are used? (p. 3).

Presenting categories derived empirically, Alkin, Daillak, and White (1979) provide eight categories in their analytic framework of local evaluation use: preexisting evaluation bounds; orientation of the users; evaluator's approach; evaluator credibility; organizational factors; extraorganizational factors; information content and reporting; and administrator style (p. 235).

Several conceptualizations of types of use have also been offered. For example, Fullan (1979) has suggested that information may result in changes in values, in understanding, in roles, in organization, or in materials. Some authors use different terms for types of use that apparently involve the same processes. As Weiner, Rubin, and Sachse (1977) observe, "These categories are neither mutually exclusive nor exhaustive" (p. 12).

Instrumental use, also called allocative use (Weiner, Rubin, and Sachse, 1977), represents the traditional, mainstream view of the use of evaluative information where the information results directly in programmatic changes. As noted earlier, this type of use rarely occurs, although some examples of instrumental use can certainly be identified (e.g., the Rockland case study in Alkin, Daillak, and White, 1979). Pelz (1978) reminds us that

". . . it is seldom possible to trace a single decision to a single piece of knowledge in a one-to-one matching between input and output" (p. 354), and Weiss (1979) concurs:

> To limit our attention to direct and immediate application of evaluation results to decisions forecloses the opportunity to understand how evaluations in fact affect program operation (p. 7).

A second, more common form of use is labeled conceptual use, i.e. use "influencing a policymaker's thinking about an issue without putting information to any specific documentable use" (Rich, 1977, p. 200; see also Caplan et al., 1975). Rein and White (1975) suggest that over time, summative evaluations change people's ideas of what can or can't be accomplished in programs. The cumulative effect of conceptual use and a variety of other information then determines future programs. The importance of this type of use is noted by Cronbach et al. (1980) who write that "stimulating a discussion that leads to gradual change in prevailing views is very likely the most important effect of evaluation research" (p. 193). They even go so far as to say that "only a minority of evaluations are to be judged primarily by their service to whoever commissioned them" (p. 61), adding later that an evaluative study "is to be judged primarily by its contribution to public thinking and to the quality of service provided subsequent to the evaluation" (p.64). This type of use has also been labeled appreciative (Weiner, Rubin, and Sachse, 1977) and "enlightenment" (Cronbach et al., 1980).

The distinction between instrumental and conceptual use may be clearer in theory than in practice. Leviton and Hughes (1979)

have written that "it is difficult to determine where conceptual use ends and instrumental use begins" (p. 10), pointing to the need for documentation of instrumental use; an example of instrumental use may be mislabeled because of missing documentation. Weiss (1979) notes that the use of research and evaluation results can more sensibly be placed on a continuum, with actual, immediate (i.e. instrumental) use at one end and the conceptual, diffuse contribution to understanding at the other (p. 10). For example, any given evaluation will have certain effects on the program it studies--textbooks may be changed, objectives rewritten, services expanded, and so on. At the same time, the results of that study contribute to an overall picture of education in the decision-maker's mind, one of many bits of such information integrated yearly.

A third type of use has been variously called symbolic (Pelz, 1978; Young and Comptois, 1979), persuasive (Leviton and Hughes, 1979), or ritualistic (Braskamp, 1980) use. Although these labels point to slightly different types of use, the distinguishing characteristic of symbolic use as an umbrella term is the pre-eminent influence of the evaluation's context on the use; the success of such use is independent of the message brought to the decision-maker (Weiner, Rubin, and Sachse, 1977, p. 13). Symbolic use can take several forms. For example, program personnel who solicit evaluations only to satisfy external funding agencies are engaging in symbolic use; the evaluation is only used to persuade the agency that the game is being played according to the rules. In persuasive use, decision-makers "[draw] on

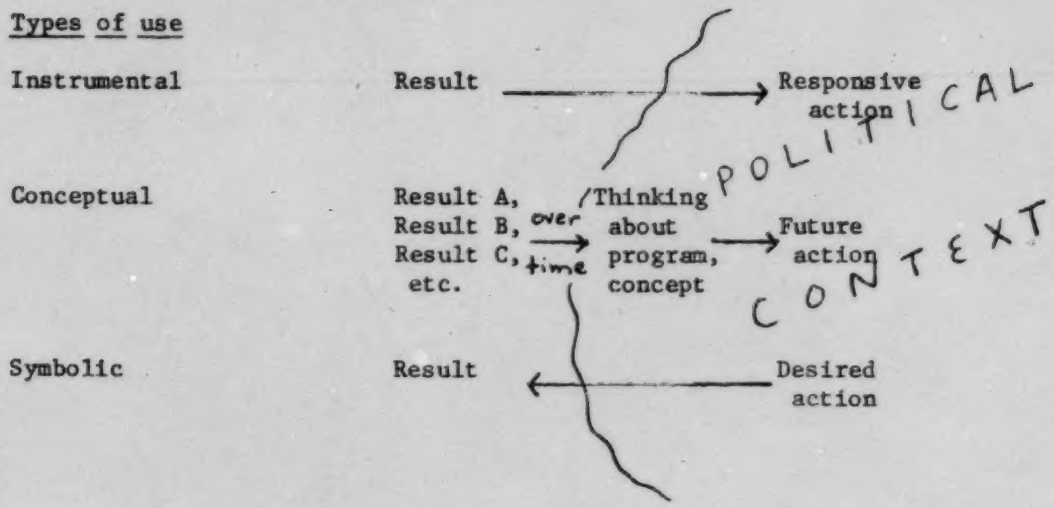
evaluation evidence in attempts to convince others to support a political position, or to defend such a position from attack" (Leviton and Hughes, 1979, p. 7). Knorr (1977) indicates another symbolic use of evaluations; she writes that Austrian administrators reported they had used evaluations to legitimize decisions (pp. 171-172). Further examples of symbolic use include gaining recognition for a successful program or discrediting a disliked program (Weiss, 1977); and as window dressing or as part of a public relations campaign (Alkin, 1975).

To label symbolic use ritualistic is a misnomer since the use can have deliberate and important program impacts. Cronbach et al. (1980) point out that "sometimes an evaluation or monitoring procedure has only a ritual or symbolic purpose. To use these labels scornfully is to misperceive the utility of symbolic acts" (pp. 159-160). Such scorn also ignores the political realities of most evaluation settings. Figure 2 diagrams the relation of these modes of use to the evaluation context. In cases of instrumental and conceptual use, evaluation results are applied to the political context in immediate or longterm ways. In contrast, symbolic use comes <u>from</u> the political context; knowing what they want to do, decision-makers turn evaluation results to their own ends, whether appropriate or not. To do this is not necessarily unethical or manipulative. It may be a method for survival. Patton (1978) has written,

> The potential for enhancing utilization lies less in its capability for rationalizing decisionmaking than in its capacity to empower the users of evaluation information (p. 35).

The symbolic use of evaluation recognizes this.

Figure 2. Types of use and related actions



The conceptualizations of use presented above may accurately reflect real world evaluation dynamics, but the shift from the more traditional view of use does produce some problems. An expanded view of use makes it more difficult to study use phenomena. For example, "It is literally impossible 'to prove' (conceptual) use" (Fullan, 1980, p. 44), and few decision-makers may be willing to recognize or own up to symbolic use. Nevertheless, the trade-off of measurability in return for a more realistic perspective on use is clearly worthwhile. As Braskamp and Brown (1980) have argued, "although the expanded definition makes utilization less dramatic and more difficult to explicitly measure and demonstrate, it represents a view of evaluation in which the role of human interaction in the communication process is given more credance" (p. viii).

Empirical Studies of Local Evaluation Use

Dickey (1981) expresses the consensus in the literature when she writes, "There have been many articles on under-utilization, but few empirical studies suggesting what variables might account for it. Supporting data are limited, and most of our 'theories' come more from common sense than from research" (p. 65). In 1975, Davis and Salasin reported that a "review of 1,200 references on [knowledge] utilization contained only 2 1/2% which pertained to evaluation, . . . even in the broadest sense" (p. 626). Conditions were similar in 1979 when Alkin and Daillak wrote that

> While much has been said and written about the problems besetting evaluation and about the underutilization of evaluation information, very few empirical studies of evaluation utilization have been conducted. Most of the literature is anecdotal in form (p. 41).

Shapiro (1979) agreed: "The literature on utilization, both applied and theoretical, tends to be ad hoc and nonrigorous" (p. 1). Cook (1978) has suggested that "the quality and imaginativeness of most (but not all) utilization studies leaves something to be desired" (p. 14).

The difficulties of conducting research on evaluation use are many. Most researchers would agree that

> The researcher who truly wishes to understand the "why?" of utilization cannot treat evaluation as a black box with inputs (characteristics, factors, etc.) and outputs (decisions), but must open up the evaluation black box and carefully study the interactions of people and events which produce the multiple consequences of evaluation and which give these consequences meaning (Alkin, Daillak, and White, p. 32).

But, even apart from the conceptual issues discussed above, practical constraints on studying the use of evaluation information abound. Leviton and Hughes (1979) note that it is often difficult to document use or to demonstrate that change at any level is due primarily to the results of an evaluation (p. 15). Decision-makers and their local school personnel may be reluctant to share their use--or non-use--of evaluation results. Stevenson (1979) writes that "verbal acceptance of findings may not be followed by appropriate action. Verbal rejection of findings may be followed by actions which imply acceptance" (p. 3). A further problem stems from the longterm effect of results; a study completed one year may continue to affect programs years after its final report is released (Holley, 1980b, p. 106).

Weiss (1979) gives five cases in which the interaction of evaluation results with a program and decision-maker leave the evaluation researcher in a quandry: 1) a decision-maker given an

evaluation study without definitive answers; 2) a decision-maker given two conflicting evaluation studies; 3) a decision-maker managing a program with no real problems or decision points; 4) a decision-maker considering many factors, only one of which is the evaluation information; 5) a decision-maker using evaluation results symbolically to justify what was pre-planned (pp. 3-6). The question of what appropriate "use" looks like in these cases is a difficult one to answer.

Despite these difficulties, a small body of empirical research on the local use of evaluation results does exist. Because true experiments on this subject are virtually impossible, the existing studies have been of four types: interviews (Alkin et al, 1974; Andrews, 1979; David, 1981; Dickey, 1979; Kennedy, Apling, and Neumann, 1980; Patton et al. 1977; Williams and Bank, 1980); surveys (Caulley and Smith, 1978; Goldberg, 1978; Lyon et al., 1978; King and Thompson, 1981); report simulation studies (Granville, 1977; Newman, Brown, and Braskamp, 1980; Thompson, 1981; Thompson and King, 1981); and naturalistic case studies (Alkin, Daillak, and White, 1979). These studies have addressed three major questions:

1. Are evaluation results used?
2. What are the characteristics of local evaluation units and of local decision-makers?
3. What factors affect use at the local level?

Tentative answers compiled from the literature provide some insight into the local use of evaluation results.

1. Given the sense in the field that results were not being used, much early research tried first to determine if results were

being used and, if so, in what ways. Some of these studies have been mentioned above in discussing the changing notion of evaluation use. The good news is that evaluation results are used. For example, although they studied the use of federal health evaluations, the findings of Patton et al. (1977) have been widely cited as providing support for the use of evaluations. In open-ended interviews with three informants per project, i.e. the project officer, the decision-maker, and the evaluator, they found that

> . . . evaluation research is used by decisionmakers but not in the clear-cut and organization-shaking ways that social scientists sometimes believe research should be used (p. 144).

Studying the local uses of Title I evaluations at 15 sites, David (1981) found that the results were used, but primarily "to meet legal requirements, to provide feedback, and to provide gross indicators of program effectiveness" (p. 31). As she summarizes, "Title I evaluations do not seem to serve as primary purposes, either as a basis on which to guide the program or as a guide to program improvement" (p. 31). The use of evaluation results and other information are used is implicit in Kennedy, Apling, and Neumann (1980) where the results of 345 interviews and the observations of roughly two dozen meetings are organized into four categories of issues: district-wide, program-level, building-specific, and clinical. Varying uses are discussed for each issue category.

The perception of local users also supports the claim that results are used. Dickey (1981) conducted a retrospective interview study of 47 Title IVC projects using analyses of evaluation reports and archival data as well as questionnaires and some telephone

interviews of local project directors. As mentioned above, over two-thirds of these users reported finding the evaluations "very" or "quite useful" (p. 71), suggesting that they perceive the results being used. In a nationwide survey of LEA users, King and Thompson (1981) similarly found that over half--60%--reported finding the results of program evaluations either "very useful" or "useful" (p. 7).

The seminal case studies reported in Alkin, Daillak, and White (1979) further suggest that evaluation results are used and in a variety of ways (see, for example, pp. 223-226). These naturalistic studies marked an important methodological addition to evaluation use research in that they involved on-site observation and interviews. The question, "Are evaluation results used," has been answered empirically, and the answer is yes.

2. The second question addressed by empirical research has been that of describing the characteristics of the participants in local evaluation use, i.e. the evaluation units and the local decision-makers. Two nationwide surveys of large city evaluation units have been conducted (Lyon et al., 1978; Webster and Stufflebeam, 1978). The portrait that emerges of these evaluation units--and it should be noted that little is known about evaluation in small districts, except that they are unlikely to have an evaluation unit (Lyon et al., p. 8)--provides little to comfort the evaluation community.

According to survey results from over 200 large districts (Lyon et al., 1978), evaluation is an in-house activity. School district personnel, rather than external consultants, do the major share of

evaluation office work, and, contrary to popular belief, these units are primarily supported by local funds, not federal or state monies (p. 57). Although little agreement exists on what constitutes basic evaluation practice or on what evaluation activities deserve priority (p. 90), evaluation in these units frequently means achievement testing: roughly 75% say that "student achievement is the dominant topic of data collection" (p. 76); and an equal percentage say that "testing is their major method of data collection" (p. 79). The two activities ranked as consuming the most time--the assessment of instructional programs and of student achievement of objectives--suggest a narrow view of evaluation (p. 83). Furthermore, data collection typically receives more attention than data analysis (p. 77), and only 15% of responding offices reported that the use of evaluation results to modify programs was either first, second, or third among time-consuming activities (p. 86).

The relation of evaluation to improved instruction is then tenuous, especially given the organizational position of many units. Evaluation units are more likely to be in one of the typical lines of authority (e.g. Instruction or Administration) than to report directly to the superintendent, but, even so, 62% of the offices are <u>not</u> located in the Instructional line. Development activities in these offices generally center on tests and evaluation instruments, rather than on instructional programs and products. This is partly explained by looking at the clientele of the average unit; roughly 60% of the time is spent with administrators compared to only 40% with instructional clients (p. 102). Also, the career patterns of many evaluators may de-emphasize instructional management experiences: a

full 42% of evaluators responding have not taught (p. 66).

The results of Webster and Stufflebeam's survey of 35 large urban school districts (1978) provide additional details of this portrait. If dollars invested suggest commitment to an activity, evaluation functions are not strongly supported in these districts; the "total evaluation bill [for 1977-1978] was $13,002,049, a mere fifteen-hundredths of one percent of the education budget" (p. 26). As the Lyon et al. study found, the evaluation departments reported putting "most efforts into testing, product evaluation, and data processing" (p. 29), rather than into program improvement. Because all evaluators are influenced by the politics surrounding them (p. 34), evaluation department heads reported having to "[sell] evaluation activities to decision-makers, in many instances, convincing them that they need information to make better decisions" (p. 29). Even in districts with evaluation units, then, the results of these surveys suggest that the use of evaluation results at the local level is a less than likely prospect.

In addition to these surveys, Kennedy, Apling, and Neumann's (1980) interviews of evaluators in 18 districts across the country provide further insight into the local evaluation office to support this view. The evaluators interviewed discussed five issues. First, the multiple clients they serve in a local setting make it impossible to focus on or please everyone all of the time. Second, the conflict between objectivity and involvement forces evaluators to create a delicate balance between independence and close working relationships with program managers. Third, the concern for professional credibility means they must aim for technical quality in difficult

working environments, where clients frequently question the educational (but not the technical) quality of their work. Fourth, to promote information use, evaluators need to employ several strategies such as engaging in early negotiations with potential users and securing the support of board policies that mandate evaluation use. Fifth, the relationship between evaluator and client was reported to influence use only when the evaluator became a personal consultant, an extremely time-consuming task, thus suggesting that it "may not be reasonable . . . to place too much responsibility upon the evaluator for improving evaluation use" (p. 118).

The characteristics of local decision-makers derived from empirical studies provide additional information pertaining to local use. As noted above, these users frequently report finding report results useful. Alkin, Kosecoff, Fitz-Gibbon, and Seligman (1974) studied the relation of evaluation reports and decision-makers in 42 Title VII projects. The methodology consisted of data collection through the analyses of evaluation reports and audits, retrospective questionnaires or telephone interviews with project personnel, and federal monitors' ratings, followed by extensive statistical analysis. In computing a local decision-making factor, they found that "when [local] project directors reported the evaluation useful in one area (e.g. in preparing reports, changing personnel, recommending project changes, etc.), they generally reported it useful in all areas" (p. 53). Furthermore,

> Whereas evidence of evaluation sophistication and comprehensiveness . . . were associated with federal decision making, local decision makers were influenced by the physical amount of evaluation information produced and general judgments of project quality. With respect to local decision making, the rule seemed to be: the more

> evaluation the better and the more positive the assessment the better (p. 65, emphasis in the original).

The image of the local evaluation user derived from this study is someone eager for information. It should be noted, however, that the study also found no recorded failures in the evaluation reports (p. 17); the possibility that local decision-makers were finding useful positive documentation of their programs might easily explain why the evaluations were appealing.

In the study cited earlier, Dickey (1981) found that

> . . . the project director is most likely to use the evaluation when he or she is interested in and committed to it, when he or she sees its procedures as appropriate and its recommendations as helpful, and when the final report is produced on time for the project's needs (p. 73).

She also found that directors of validated projects were more likely to label evaluation information useful. This conflicts with Granville's finding (1977) that, in a simulation study, principals responding to a report on an innovative program were more likely to incorporate objective evidence into their decisions than those reading a report on a "routine" program (p. 6). These results may clearly be due to the numerous methodological differences between the studies, but the variable of program innovativeness needs additional clarification.

The four types of issues raised by local users interviewed by Kennedy, Apling, and Neumann (1980) are presented in Table 1 and suggest the variety of concerns these multiple clients face in using evaluation and testing information. The problems and information needs of local users therefore differ radically. Smith (1980b) found that school board members had a different perspective toward

Table 1. Issues raised by users in Kennedy, Apling, and Neumann (1980)

Issue Category	Topics of Concern
District-wide	a. Federal influences b. Changing enrollments c. Student behavior d. Testing
Program level	a. Responsibility and authority b. Compliance c. Funding
Building-specific	a. Implementing district policies b. Building management c. Program improvement
Clinical	a. Students' instructional needs b. Students' social and emotional needs c. Referral of students to special programs

accreditation studies than administrators; the users who provided the lengthy set of issues in Table 1 similarly must view evaluation from differing perspectives.

Three studies have examined administrators' and project staff's attitudes toward evaluation and evaluators. David (1981) discovered two underlying attitudes toward evaluation among the Title I local staff: a "narrow and usually negative" perception of evaluation; and the "'true believer syndrome,'" i.e. the feeling that, as one director said, "'We are successful even if we can't show it on paper'" (p. 38). The survey described in King and Thompson (1981) found that only 28% of the LEA users responding reported that the program effects they most care about can be directly measured; well over two-thirds (72%) felt that these effects can only be measured indirectly (43%) or not at all (29%) (p. 8). Users who feel that key program effects cannot be measured may find working with evaluators a frustrating experience, although it must also be noted that most of those responding reported infrequent contact with evaluators (p. 9).

Both this survey and a related simulation study have documented that administrators perceive the characteristics of different evaluators. In the survey, administrators were asked to categorize a typical evaluator in their district according to his or her technical skill (good or below average) and political skill (good or below average). The four combinations of these variables create a typology of evaluators first suggested by Meltsner (1976). The local users labeled their typical evaluator as follows:

Good in technical skill, good in political skill (entrepreneur)	49%
Below average in technical skill, good in political	26%

skill (politician)

Good in technical skill, below average in political skill (technician) 17%

Below average in technical skill, below average in political skill (placeholder) 8%

The recognition of evaluators' political skills in LEA settings is evident in that 75% of the respondents labeled their most common evaluators as high in political skills. The relatively equal importance of technical skills is reflected in the 60% rating for those categories with high technical skills. These results contrast in part with the findings of Thompson (1980) in which the evaluators in an urban district evaluation unit were identified as either entrepreneurs or placeholders (p. 63). However, the second study used a factor analysis of evaluators' reactions to statements to label them, and the difference in subjects and method may well account for the difference in the results.

An additional study by Thompson and King (1981) using simulated evaluation reports found that administrators may implicitly recognize differences among evaluators when they read reports. Given reports that varied in the use of technical and political language, principals responded differently. The results suggested, however, that administrators were more attentive to the technical merit of reports than they were to the evaluator's political sensitivity. Although the results of any simulation study must be considered tenuous, these findings, coupled with the work discussed above, suggest that users do perceive differences among evaluators. To the extent that perceptions affect the subsequent use of results, they may be important variables to consider.

As the results described previously suggest, the answer to the

second question--what are the characteristics of local evaluation units and of local decision-makers?--is highly complex. There are, however, at least two generalizations to be made. First, local evaluation units are typically not involved in program improvement; and second, although some research supports the claim that local users find evaluation information helpful, other data suggest that these users may feel that evaluation is not of value to them. The picture that emerges of the local evaluation context, then, supports it as a setting for potential non-use.

3. The third question is, finally, the key to improving the use of evaluation information at the local level. By knowing the factors that affect use, evaluators and decision-makers together can work toward increasing the appropriate use of results. The discussion of what empirical research has learned about these factors will be separated into two sections, presenting first those factors within the evaluator's control and second those factors pertinent to a local evaluation context, but out of any one person's control.

Factors Evaluators Can Control. Two variables that have received considerable discussion in the evaluation community are the technical quality of evaluation studies and of evaluation reports. The Joint Standards contain both methdological standards (D3, "Described Purposes and Procedures"; D5, "Valid Measurement"; D6, "Reliable Measurement"; D8, "Analysis of Quantitative Information"; etc.) and explicit standards for reports (A5, "Report Clarity"; A6, "Report Dissemination"; A7, "Report Timeliness"). Because the need for the highest quality

evaluation methodology and reports is intuitively obvious, little empirical work has been conducted on these topics. However, as Weiner, Rubin, and Sachse (1977) point out, to say that evaluation results are underused because of their low technical and scientific quality is to display a "faith in rationality," rather than an awareness of "evidence concerning the factors influencing the utilization of evaluative information" (p. 4).

The results of three studies suggest that the use of evaluation information is not directly related to the methodological quality of the evaluation itself. Patton et al. (1977) write that ". . . there is little in our data to suggest that improving methodological quality in and of itself will have effect on increasing the utilization of evaluation research" (p. 151, emphasis in original). Dickey (1981) similarly found that methdological sophistication (e.g. hypothesis-testing statistics, multiple data sources, control/companion groups) had no relationship to the level of use (p. 73). As mentioned above, while the evaluators interviewed in Kennedy, Apling, and Neumann (1980) worried about the methodological quality of their work, the users of their results were more concerned with the "educational quality" than with the technical quality (p. 111). Webster and Stufflebeam (1978) reported that the evaluation departments of the districts they surveyed did have highly trained, methodologically competent evaluators (p. 32). Evidence suggests, however, that improving the technical quality of results will not necessarily improve the level of their use. Not being methdological experts, decision-makers will typically use or not

use evaluation results for reasons other than their technical merits.

The evaluation report is a second factor affecting use that evaluators can control. Several studies have examined aspects of the effect of reports on use. A series of studies by Braskamp, Brown, and Newman (see, for example, Newman, Brown, and Braskamp, 1980; Braskamp, Brown, and Newman, 1981) has systematically investigated the "relationship among the characteristics of an evaluator, an evaluation report, evaluation audience characteristics, and audience responses" (Newman, Brown, and Braskamp, 1980, p. 30), using communication theory as a conceptual framework. Their results suggest that several report variables affect the use of evaluation information.

a. The message source--the evaluator--"generally affects an audience's reactions not only of their ratings of the message source (evaluator) but also affects their extent of agreement with the evaluator's recommendations" (Braskamp, Brown, and Newman, 1981, p. 6). Both the title and sex of the evaluator can affect audience reactions. Even though readers read identical reports, they rated those they thought written by a "researcher" as significantly more objective than those written by an "evaluator" or a "content [art] specialist." Report readers were less likely to agree with reports which they thought were written by female evaluators when the field differed from their own; in their own field, they were less critical of the female's results, although they were still more likely to agree with the male evaluator's recommendations.

b. The message content can also affect readers' reactions. The use of jargon and data can affect audience ratings of technicality and difficulty. In one study, the report rated most difficult by readers included jargon, but no data support. "Generally, reports containing both jargon and data were rated more useful regardless of whether the readers were professionals in the field or were lay persons from another field" (Newman, Brown, and Braskamp, 1980, p. 33). Rated next highest were reports lacking both jargon and data use, suggesting an interaction between the use or non-use of jargon and data in the reactions of readers.

A study by Brzezinski (1979) rated evaluation reports using the criteria outlined in the Joint Standards. The studies ranked the highest were "those that were the longest and most research report-sounding" (p. 6). This suggests that the Standards criteria, which give high points for appropriate technical language and data, may then lead to reports that users find most useful. To the extent that local users also want as much information as possible (Alkin et al., 1974), these technically sound reports may also be the most effective.

A word of caution should be added here, however. The variables discussed--report use of technical language and data, and report length-- have not been exhaustively studied. Other variables, for example, program innovativeness or the inclusion of negative results, may well affect local use of evaluation information. The finding of Brickell et al. (1974) concerning the length of desirable reports in the federal bureaucracy may be

equally true at the local level:

> It was not unusual for officials to request short reports for themselves and longer ones for their subordinates, but then to find that their subordinates when interviewed wanted short reports for their own use and suggested longer reports for their subordinates and so on down the hierarchy (p. 60).

Further, other research supports the view that administrators prefer qualitative information over quantitative information. Alkin's (1980b) well-known naturalistic use studies yielded the conclusion that "little evidence was found in the case studies that research rigor was an important factor affecting utilization" (p. 24). Simulation research by Brown and Newman (in press) is even more dramatic:

> The simple addition of an inferential statement, such as "these differences were statistically significant at the .05 level," however, resulted in lower levels of agreement (with policy recommendations). In fact, for three of four recommendations, the inclusion of the inferential statement resulted in levels of agreement lower than in the no data [experimental] condition.

However, it is important to note that the use of data does interact with other result features in determining audience reaction (Brown, Newman, and Rivers, 1980, p. 72), so a simple interpretation of these results is not possible.

The effect of report timeliness seems, like technical quality, intuitively obvious; results that are unavailable to a decision-maker at the time of a decision cannot be used. Randall (1969) graphically portrays this situation:

> There is a timeworn and oft-recurring spectacle of the frantic but finally productive researcher-evaluator, who rushed into the executive offices with his data analysis in hand, only to find that the executives, several months previously, had made the important decisions that locked up the monies and committed the organization for the ensuing months ahead (p. 1).

The results of Dickey (1981, p. 73) provide support for this assumption. However, two studies provide evidence to the contrary. Patton et al. (1977) reported that timeliness was not a critical factor in the use of federal health evaluation results (p. 150), and Kennedy, Apling, and Neumann (1980), studying local users, also labeled timeliness of questionable importance:

> Although we found several major decisions in the making, the importance of timely data was hard to gauge. . . [E]ven when decisions were based on formal studies, timeliness was not critical (p. 131).

They note elsewhere that for project directors, the results of formative evaluations are not available quickly enough; action must be taken using available data, even if it is last year's summative study. The case studies in Alkin, Daillak, and White (1979) also demonstrate the problems of reports that are made available only after the next year's planning cycle.

The effect of report timeliness is, then, uncertain and perhaps even, in Cronbach et al.'s opinion, "much-overrated" (p. 63). Clearly, decision-makers in local settings use information they are interested in as soon as the data are available in any form; as Patton (1978) notes, the formal report of an evaluation will hold no surprises for an alert decision-maker (p. 205). To suggest that report timeliness is an important factor is to suggest a form of instrumental use. But as was noted above, the instrumental use of evaluation information is, in general, limited. Although timeliness may be important for short-term instrumental use, it may not be as important for the more general use of evaluation results (Young and Comptois, 1979).

c. The characteristics of the receiver can also affect the use of information. Ratings of the usefulness of the results of an external evaluation differed depending on the organizational position of the audience. Other important variables were the level of professional experience and the field of the reader. Involved here, too, is the audience's perceived need for evaluative information in a particular area (Brown, Newman, and Rivers, 1980). "Those with a high perceived need for evaluation agreed more with the evaluator and were generally more satisfied with the information they had available than those with a low perceived need" (1981, pp. 11-12).

A finding of King and Thompson (1981) suggests the importance of the evaluation report to the local user. Almost two-thirds of both principals and superintendents ranked written evaluation reports "which identify difficulties and discuss possible actions to correct them" as the most important form of evaluation reporting in their systems. To the extent that evaluation reports can be improved in meaningful ways (e.g., including appropriate data, gearing reports to certain organizational positions), empirical evidence suggests that the use of results may then also improve.

Factors Related to the Local Evaluation Context. While the methodological quality of evaluations and the evaluation report are factors that local evaluators can work to control, other factors affecting use flow from the context of evaluation. The factors in this category that have been empirically investigated can be divided into two topics: the issues local users face and

the process of local evaluation use.

The interviews summarized in Kennedy, Apling, and Neumann (1980) mark an important addition to the literature because they provide evidence that the non-use of information may in many cases be beyond the evaluator's control. The four categories of issues the users described--district-wide, program-level, building-specific, and clinical--suggest the variety of ways information is used at the local level. The use of evaluation information is discussed in all but the clinical chapter, which deals only with the use of test results. As seen in Table 1, the types of issues faced at different organizational positions vary radically. District-wide issues are reactive, rather than pro-active, created by the changing conditions the institution faces. The issues program directors face often require increased personal advocacy; in such cases, personal information can often obviate the need for more formally collected data. Building-specific issues involve building management where little systematic information is used in implementing district policies.

The summary paragraphs at the end of the three chapters describing these issues are thought-provoking.

Following Chapter 3, "District-Wide Issues":

> These findings raise some interesting questions about the use of evaluation and test data. . . [O]ne wonders whether an evaluator (or anyone else) could ever predict what the issues would be in sufficient time to generate data that would specifically address . . . [the unanticipated] issues [that characterize district-wide concerns]. . . If [the currently available descriptive statistics are appropriate to district-wide issues]. . ., then the role of evaluation studies per se--that is, assessments of the merits of current practices--may be doomed to be a small one, for questions of merit seem to take a back seat to questions of new needs that must be addressed (p. 45).

Following Chapter 4, "Program Issues":

> The situation program directors find themselves in raises some important questions about the role that evaluation and testing can play in program improvements. If, for example, we accept the fact that program managers, out of organizational necessity, must use data for advocacy, should we also accept the fact that such uses will probably also entail distortion of the findings? How does one draw the line between appropriate and inappropriate uses of data in advocacy situations? Furthermore, data tend to be used in group settings, that is, a program director discusses its implications with one or two trusted staff members or an evaluator and then uses it to persuade others. The process has many layers and is diffuse. How would one intervene to improve it? (p. 65).

Following Chapter 5, "Building-Specific Issues":

> These findings are significant in that a considerable body of literature suggests that building principals are key contributors to the quality of education in their schools. Yet, for the most part, the issues they face are not issues that could be identified, clarified, or solved via systematic data (p. 82).

In other words, the use of evaluation information at the local level may be severely limited because of the kinds of issues users face at this level. The question for evaluators then becomes one of determining those issues and situations that <u>can</u> benefit from data collected systematically and of working to insure that the information is presented in useful ways. To do this effectively will require a thorough understanding of the process of evaluation use at the local level, the second set of use factors related to the context of evaluation.

The many variables affecting the process of local information use have not been documented empirically. In their study, Patton et al. (1977) found only two factors significantly related to the use of evaluation results: a "political

considerations factor" and the "personal factor" (p. 149), i.e. the presence of at least one person in an evaluation study who cared about the process and using its results. In an LEA setting, Granville's (1977) simulation study similarly found that the "political" and "social influence" factors significantly affected principals' decisions. Granville's social influence factor indicated the importance of the professional peer group at the local level. These results suggest the obvious: to work in a school district is to live in a political environment where specific individuals can make a difference and can actually make evaluation use occur.

The conclusions of empirical studies have repeatedly recommended a collaborative approach to local evaluation efforts, i.e. evaluators and decision-makers working together to create and use evaluation information. Both the political and personal factors suggest the validity of such an approach. Based on the study of Title IVC project directors cited earlier, Dickey (1981) concluded that "evaluators should adopt a more collaborative role, involving the decision maker and the staff in decisions about the evaluation" (p. 76, emphasis in original). She also writes that "evaluation, conducted at a distance, is less likely to be valid and unbiased and thus, less useful" (p. 76).

Goldberg (1978) found that New York City ESEA administrators received helpful information from evaluations in the problem awareness step of decision-making, but not in the later steps, i.e. in finding possible courses of action or in choosing among alternative courses of action (p. 19). He

recommended collaboration as a way to increase the use of evaluation information throughout the decision-making process.

Two case studies of Alkin, Daillak, and White (1979) provide graphic evidence of how collaboration can help the use process. In the Bayview example, the "evaluator, assigned by chance to a project, . . . [became], in a very short period of time, a trusted advisor to decision-makers who . . . [had] little predisposition toward evaluation" (p. 208). The contrasting role of the evaluator in the Valley Vista example shows how an outside evaluator intent on collecting mandated information can force participation against the local staff's wishes--but not use.

"Collaboration" can be as straightforward as having evaluators share results with decision-makers as the evaluation progresses. David (1981), for example, states that "the provision of feedback, particularly when explained in person, provides what may be a necessary but not sufficient condition for utilization of the evaluation information" (p. 29). In their survey of evaluation offices, Lyon et al. (1978) noted that most reported feedback to project personnel consisted of "informal verbal recommendations rather than formal written ones" (p. 87). Further, "the importance of verbal, rather than written, evaluations was also underlined in fieldwork activities" (p. 43). It is possible that this degree of collaboration has helped local decision-makers to apply evaluation results in the subtle ways now recognized as constituting use.

Although research thus suggests that collaboration may be a good idea, the constraints of practice may limit this approach.

The evaluators interviewed in Kennedy, Apling, and Neumann (1980), for example, knew that extensive involvement in a program--becoming "close enough to program directors to act as their personal consultants"--would influence information use, but they knew too that it was a "role that required more time than most of them were able to spend" (p. 117). Kennedy et al. give three reasons that may limit the efforts of evaluators at encouraging use: (1) some people already use information effectively and the effect the evaluators could have had is unclear; (2) others had not been converted to use despite the presence of evaluators who "possess an imposing array of personal skills and qualities"; and (3) much of the use of data "occurred because of other, institutional pressures" beyond the single-handed control of evaluators (pp. 118-119). The role of professional knowledge in evaluation use, for example, has not yet been carefully studied.

In recording that two-thirds of the responding evaluation heads felt that their personnel resources were inadequate, Lyon et al. (1978) provide another reason for questioning the probability of the practicality of the collaborative approach. A further reason the approach may be unsuitable stems from evidence presented in Smith (1980b) that merely increasing user participation in an evaluation study may not increase the use of the recommendations (p. 64). Whether board members' involvement in accreditation studies parallels other local users' involvement, however, remains to be seen.

Like the second question, the answer to the third

question--what factors affect use at the local level?--is complex and, to date, not clear. Certain factors are known to be important in the process of local evaluation use, and research has suggested the efficacy of a collaborative approach. However, knowledge of the use process remains limited.

Strategies for Improving Use

Although few empirical studies of the local use of evaluations exist, frequent mention is made in the evaluation literature of strategies for improving such use. Cronbach et al. (1980) argue that "if any single intellectual sin is to be blamed for the present chaos, it is the readiness to make general assertions that supposedly apply to all evaluations" (p. 51). Concurring, Patton (1978) warns that

> The overall problem of underutilization of evaluation research will not be solved by compiling and following some long list of evaluation proverbs and axioms. Real world circumstances are too complex and unique to be routinely approached through the application of isolated pearls of evaluation wisdom (pp. 19-20).

The following, therefore, will be a general discussion of the factors identified in the literature as affecting use and strategies for addressing these. The discussion will again be divided into those factors that are more or less within evaluators' control and those more related to the evaluation context that cannot be controlled, but must be dealt with.

Factors Evaluators Can Control. Although methodological quality has not been shown empirically to be either necessary or sufficient for use to occur, common sense and the Joint Committee's Standards suggest that it is a minimum criterion for good evaluation practice. Alkin (1975) points out that "apparently, one of the best

conscientious defenses against non-utilization of evaluation findings is a technically sound, methodologically credible study" (p. 207); without such defenses, critics can easily dismiss evaluation results by attacking their technical credibility (p. 206). Somewhat idealistically, Leviton and Hughes (1979) suggest, "If [methodological] quality does influence use, it is likely to do so primarily through increased trust that the findings are an accurate picture of the program" (p. 25). While it is true that some administrator disdain for empiricism may be appropriate because quantitative forms of representation "inherently are insensitive to some of the significant aspects of classroom life" (Eisner, 1980, p. 11), it is also true that good methodology will insure results that are worth using. "The central message in this regard," however, "is that it is not enough to conduct methodologically sound research" (Johnson, 1978, p. 12).

In addition to sound methodology, as noted above, a second factor within the evaluator's control, is the evaluation report, in both its oral and written forms. Empirical results have demonstrated the varying effects of the message source, message content, and--perhaps--its timeliness. The importance of effective communication by evaluators has been highly emphasized in non-empirical writing. The likelihood of the evaluation report itself having an impact may be small because, as Cronbach et al. (1980) point out, "The evaluator's final, formal report is essentially an archival document, not a live communication" (p. 185). However, given the local users' interest in having such reports, evaluators must work to make them as good as possible. "Most of these reports," writes Denny

(1980), "have one omnipresent quality: they are dull, dull, dull" (p. 4). Datta (1979) reports that "Although titled 'evaluations,' more accurately . . . [the evaluation reports she examined] might be presented as descriptive statistical accounts of some aspects of educational programs" (p. 17); as such, they may not be useful to local decision-makers.

Suggestions for improving reports include using details of events and people, making the evaluation report a "well told story" (Denny, 1980, p. 5); including details of the evaluator's personality, i.e. his or her competence, belief, style, etc. (Denny, 1980, p. 6); using executive summaries on different colored paper (Alkin, 1975, p. 208); and using appendices to hold more technical information (Alkin, 1975, p. 209). Datta (1979) also recommends the executive summary approach, i.e. that evaluators "address the few important questions, present the evidence, and state the action implications in one page or less" (p. 23). Brzezinski (1979) points out, however, the limits of such summaries for truly informed decision-making (p. 1).

Should evaluation reports contain recommended actions? Empirical research has not provided a definitive answer to this question. Although the answer is clearly tied to the evaluator's conceptual approach to evaluation (see Webster and Stufflebeam, 1978), to the extent that such recommendations may assist harried decision-makers, they may well be of value. Haller (1974) observes that

> Evaluation problems concern decision, decisions presume the existence of alternatives, and so the purpose of evaluation is to help delineate alternatives and to provide information to help decision makers arrive at more rational choices (p. 403).

However, Weiss (1979) notes that many evaluators are not policy oriented; "evaluators do not always--or even often--come up with data that give explicit guidance for action" (p. 3). Zepeda (1980) suggests that this situation is not entirely satisfactory: "Informing local decision makers that the local Title-I program is not effective does not give them the information they need to improve it" (p. 1).

Will administrators perceive recommendations as an unwarranted intrusion into the policy arena? They will if the evaluator's offerings take the form of grandiose schemes. They may not if specific policy alternatives are mentioned and the evaluator merely presents objective evidence, both for and against, regarding the best predictions about likely program impacts. Evaluators sometimes perceive administrators as defensive of their territory, and in some cases these perceptions are fully justified. But many administrators do not feel threatened by good ideas and will take their wisdom wherever they can get it, especially if they believe that the recommendations are being offered in a sincere attempt to help rather than in an insincere attempt to be Machiavellian. However, as Newman and Brown's (1980) results indicate, the utility of this effort will partly be determined by the situation-specific personalities and the needs of the involved administrators.

The literature suggests that the informal contacts between evaluators and users are more important than written reports in effecting use. Four of the six themes discussed in the New Directions Sourcebook entitled Utilization of Evaluative Information (Braskamp and Brown, 1980) point to the importance of this interaction: utilization as an "immediate concern of evaluators"; the "active role

of evaluators in enhancing utilization"; the need for a "highly interactive consulting relationship between evaluators and key decision-makers"; and the "importance of the communication process" (pp. 91-94). Consider also the following quotations.

> The evidence on dissemination suggests that informal communication that cuts the red tape may enhance utilization, although quality of information may sometimes suffer and dissemination will be haphazard (Leviton & Hughes, 1979, p. 21).
>
> For while information is an essential resource for decision makers, the manner in which it is converted into policy is based as much or more on interpersonal, organizational, and psychological factors than on the actual information itself (Guskin, 1980, p. 45).
>
> Utilization is usually the result of the relationship between the evaluator and the user more than anything else. If the user knows and respects the evaluator, utilization has its highest potential (Holley, 1979, p. 8).
>
> Much of the most significant communication is informal, not all of it is deliberate, and some of the largest effects are indirect (Cronbach et al., 1980, p. 174).

The message for evaluators seems clear: informal communication with the information users must be stressed throughout the evaluation process if results are to be used. It is not always possible to anticipate when information will be needed in the service of educational decision making. In fact, as Brickell, Aslanian, and Spak (1974) note, at the federal level, the administrator "can never know when he will need it [evaluative information]. The process of government decision-making is not so orderly or regular that he can schedule his need for information" (p. 24). Kennedy, Apling, and Neumann (1980) provide evidence that the same is true of local administrators who deal with district-wide issues.

Wise (1980, p. 15) and Cronbach et al. (1980, p. 160) label the

evaluator an educator whose job it is to teach others about the program studied. The information presented must be

> . . . understood, credible, and coherent to the intended audiences and the evaluator must be a guide for politically feasible action. . . [I]f a two-way communication channel between the evaluators and users is established, provisions for trust and mutual problem solving are more likely (p. 11).

Evaluators are, then, in many ways responsible for making use happen.

Lee and Holley (1978) give advice on how to disseminate evaluation results. Included among the practical tips are such things as training the press to properly interpret evaluation data and using audio-visual aids during oral presentations. In a more theoretical paper, Holley (1980a) suggests that evaluators, knowing the political nature of evaluation settings, should even use potential use as a key criterion in the allocation of evaluation dollars. It is better, in her view, to attack difficult "mountains" that users are interested in than to work on smaller, more methodologically tidy "molehills" where results may not be used. Alkin (1975, p. 202) and Bosco (1971, p. 70) suggest that potential and intended uses should be specified early in the evaluation process.

Factors Related to the Local Evaluation Context. It is not possible finally to draw a fine line between utilization factors that evaluators can control and those they cannot. Even those that are most controllable--the methodological and report quality discussed above--can clearly be affected by the complex interactions involved in an evaluation setting. The practice of "fuddling," i.e. "allowing evaluation results to disappear in a ream of paperwork, excuses, and meaningless exchanges" (Davis and Salasin, 1975, p. 623), can limit

the effect of even the finest results. The characteristics of local programs and decision-makers are an important part of the local evaluation context, and empirical studies have suggested that certain characteristics affect the use of evaluation information. Other writings support this view.

Certain "program controlled factors" work against use in any evaluation context (Alkin, 1975, p. 207). Holley (1980b) refers to "Required Losing: Factors That Prevent the Use of Required Evaluation Reports at the Local Level." She gives two such factors: (1) the requirement that mandated reports include lists of numbers "that are in themselves devoid of meaning"; and (2) an emphasis on objectives that may be good for planning, but inappropriate for evaluation (p. 107). Grobe (1978) adds three additional factors that work against effective use: (1) the fact that local decision-makers are unfamiliar with data; (2) the high emotional involvement of local project staff; and (3) the local expectation that precise answers come out of evaluation only infrequently (pp. 2-3).

Another factor affecting use, as Granville (1977) and Dickey (1981) demonstrate, is the age, innovativeness (i.e. routine or novel), and/or maturity of a program. Brown (1973) explains,

> The question as to when and which levels of information school management needs depends on the length of time a program has been in operation and the degree to which that project overlaps other programs within the school system (p. 2).

Cronbach et al. (1980) discuss four stages of program maturity that apply equally well to local and national programs: the breadboard stage; the superrealization; the prototype; and the established program (p. 108). Depending on the program's maturity, the use of

evaluation information may differ, and strategies for improving use must consider this. "Most of the time," for example, "the fully established, mature program is allowed to go its own way" (Cronbach et al., 1980, p. 113); Kennedy, Apling, and Neumann (1980) write that at the local level, fully developed programs are rarely eliminated, but instead are simply not expanded (p. 60).

The information user's organizational position is another factor affecting the type of information and appropriate use strategy. Leviton and Hughes (1979) point to a distinction between the use of evaluation for policy decisions and for programmatic decisions (p. 12), and Brickell, Aslanian, and Spak (1974) write

> Those officials who are in a position to control the project from day to day or month to month and who are responsible for exercising such control have a far greater interest in monitoring reports than decision-makers at higher executive levels (p. 59).

Furthermore, evaluators should remember that "research is often most useful to those who do not have the authority to promote a policy, i.e., teachers" (Hamilton, 1980, p. 7).

The empirical research cited above suggests two additional factors critical to the use of evaluation information: the political considerations factor and the personal factor. Although many evaluators may still feel uncomfortable admitting it, evaluation is inherently a political activity. As Cohen (1972) writes, "To the extent that information is an instrument, basis, or excuse for changing power relationships within or among institutions, evaluation is a political activity" (p. 139). The implications of this fact are identified by Dickey (1979) when she writes

> Judgment has an awesome ring, and it is not surprising that those who are being judged feel anxious, even threatened.

> Add to this the political context in which the process takes place (and there is always a political context) and we have all the ingredients for dysfunctional communication--high levels of stress leading to communication patterns arising from individual defense mechanisms (p. 3).

Certainly the political considerations factor, whose existence is supported by empirical evidence, can help or impede the use of results. As Cronbach et al. (1980) have written, "No matter how excellent a study is technically, its facts will not sweep political sentiment and power aside" (p. 46). This does not mean that evaluators must themselves engage in political machinations. However, as Meltsner (1976) suggests, the effective evaluator "tries to understand political considerations and then to make them an integrated and explicit part of his analysis" (p. 43). According to Brown and Braskamp (1980), "This means that the relationship between the evaluator and key program staff, and the evaluator's understanding of the organization in its internal and external political environment, are critical for successful utilization" (p. 93). This suggests that evaluators must work at understanding the politics of their agencies and attempt to meet the political needs of involved persons whenever doing so will not jeopardize the integrity of the evaluation.

In addition to the political considerations factor, evaluators at the local level must take advantage of the personal factor. Leviton and Hughes (1979) point out that the impact of individuals on organizations is rarely great and suggest two reasons for the effect of the personal factor: (1) in smaller settings individuals may be able to make a difference; and/or (2) in larger bureaucracies the difficulty of effective communication may allow individuals the

opportunity to make a difference (pp. 27-28). Whatever the cause, the effects of the personal factor should not be ignored by local evaluators.

One important way to go about this is by targeting the evaluation toward identified information users (see, for example, Patton, 1978). As Johnston (1978) writes,

> There is a sort of ecology for each educational program, a network of people in different roles who influence (or are influenced by) the outcome of the program being evaluated. If this is true, and research utilization is the goal of the evaluator, then there are multiple audiences for an evlauation, not just the decision-maker who commissioned the evaluation. So the evaluator has a first task of identifying who these other actors are (p. 1).

The variety of issues raised by users in Kennedy, Apling, and Neumann (1980) certainly supports the need for this effort on the part of evaluators.

However, it was noted above in discussing the difficulties inherent in collaboration that the effort can be frustrated by the complexity of the organizational network. As Randall (1969) explains, "Typically, the decision process in an organization involves a complex network of persons who have varying degrees of influence on the one who may hae constituted authority to make any given decision" (p. 7). The situation is further complicated because, as Granville (1978) notes, an evaluation study "has to persuade not just the people who ostensibly make the decisions, but also the people <u>they</u> have to persuade" (p. 29, emphasis in original). Thus, Alkin and Kosecoff (1973) conclude that "identification of the program's decision maker[s] is perhaps the most elusive variable associated with a decision context" (p. 3).

Assuming that decision-makers are identified, non-empirical writings have strongly supported a collaborative approach as building on the strengths of both the political considerations and personal factors. Gray (1979), for example, has developed a procedure and manual for collaborative evaluation. Cronbach et al. (1980) suggest that the evaluator "should engage others in a collaborative attempt to understand social events and take appropriate action. Influence comes from engagement, not detachment" (p. 153). Patton (1978) has recommended the collaborative development of evaluation designs; "it is crucial that identified decisionmakers and information users participate in the making of measurement and methods decision so that they understand the strengths and weaknesses of the data--and so that they believe in the data (p. 202). Ross (1980) has even suggested that evaluators and administrators should together specify decision rules in advance of program implementation, rules specifying what decisions will be taken if various evaluation results occur (p. 66).

Certain practical difficulties with the collaborative approach have been discussed previously. Two others are suggested in the literature. First, what Patton (1978) has termed the "goals shuffle" can readily destroy efforts to estimate discrepancies between program goals and actual program outcomes:

> The goals clarification shuffle involves a sudden change in goals and priorities after the evaluator is firmly committed to a certain set of measuring instruments and to a research design. The choreography for this technique is quite simple. The top priority program goal is moved two spaces to either the right or left and four spaces backward (p. 100).

Second, Rossi (1972) has argued that what might be called "methodology shuffles" can also occur if evaluation results prove unpopular:

> It is easy to attack the methodology of any study: methodological unsophisticates suddenly become experts in sampling, questionnaire construction, experimental design, and statistical analysis, or borrow experts for the occasion (p. 229, emphasis in original).

Collaboration with local decision-makers runs the risk of subversive activities.

Can a collaborative approach to evaluation work in local settings? The answer to this question is a hopeful yes--maybe. As explained above, the evaluation literature has increasingly recognized that the process of change in organizations is rarely abrupt or dramatic; and the instrumental use of evaluation results is therefore an unrealistic expectation. Evaluation's strength comes from its ability to assist in incremental accommodation; to provide needed information to decision-makers; to, in Patton's words, reduce their uncertainties about the program that surrounds them (1978, p. 31).

Weiss (1972) writes that the "use of evaluation appears to be easiest when implementation implies only moderate alterations in procedure, staff deployment, or costs, or where few interests are threatened" (p. 320). In this light, a collaborative approach between evaluators and potential users makes theoretical sense, the evaluator assuming the role of change agent (Davis and Salasin, 1975; Joint Committee, Standard A8, p. 47) or linking agent (Havelock, 1968; Hayman, 1979; Bank, Snidman, and Pitts, 1979). "Assisting in gradual accommodation, evaluation is both conservative and committed to change" (Cronbach et al., 1980, p. 157). At least part of the appeal of this approach is its minimal requirements. Major organizational changes (see, for example, Weiner, Rubin, and Sachse, 1977, for a proposed alteration of the federal evaluation structure) are unlikely

to occur in local education agencies; in some cases collaboration no doubt already takes place; in others, all that is required will be a shift of emphasis.

The evaluation literature contains numerous suggestions for making collaboration successful. Certainly the informal reporting procedures discussed above will encourage the type of communication essential for working together. Targeting evaluations and identifying appropriate issues will also help the collaboration process. Alkin and Daillak (1979) argue that "evaluators who concentrate on the mandated evaluation tasks run the very real risks of losing the local audiences" (p. 47). This suggests that evaluators should concentrate evaluation efforts on the highest priority information needs of specific administrators, even if these needs require work beyond that mandated by external funding agencies. As Alkin, Daillak, and White (1979) note, "If the evaluation addresses a pressing concern of a potential user, then the evaluation information is more likely to draw, and hold, the user's attention" (p. 238). Patton (1978) goes so far as to suggest enhancing use "by focusing on fulfilling one purpose extremely well, so that at least the decisionmakers' central questions are answered (p. 83).

Since administrators are not always able to anticipate or articulate future information needs, evaluators "should anticipate questions and be proactive" (Law, 1980, p. 74). Stake (1973) makes a similar point.

> The evaluator, I think, has a responsibility to snoop around and to guess at what decisions may be forthcoming. He should use these guesses to orient his evaluation plan (p. 305).

Gorham (1970) argues that evaluators should "be clairvoyant about forthcoming issues" (p. 104). Cronbach et al. (1980) suggest that ideally the evaluator "interprets outcomes broadly and envisions possible negative effects as well as the intended positive ones" (p. 260). In short, evaluators should identify some evaluation issues on the basis of emphatic and proactive anticipation of administrators' future information needs. If these anticipated needs do not arise, evaluators need not highlight the results of the inquiries that they initiated. In any case, evaluator credibility should be improved when administrators sense a sincere effort to be responsive, even if this anticipation is not always precisely accurate.

To be effective collaborators, evaluators must also demonstrate to project personnel and administrators that they sincerely care about the needs of program staff and the program's clients. This is the other side of the personal factor and means that evaluators must try to not be threatening and authoritative; how evaluators comport themselves affects the psychological frameworks with which administrators interpret evaluative information. This conclusion may discomfort some evaluators who believe they offer objective truth that has intrinsic value independent of evaluator personality or approach. Nevertheless, administrators have their own paradigms for viewing the world--paradigms that are rational in their context--and evaluators must accept that the manner in which the evaluator interacts with administrators and staff with affect, in an important sense, the credibility of subsequent evaluation results for administrators. One strategy for avoiding conflicts is to prepare a formal memorandum of agreement prior to a proposed evaluation in which purposes and

procedures are clarified (Webster and Stufflebeam, 1978, p. 39). Another is to be "active, reactive, and adaptive" (Patton, 1978).

Evaluability assessment is one procedure with great potential for local evaluation efforts. Also known as exploratory evaluation or accountability assessment (Rutman, 1980, p. 12), evaluability assessment is used to determine in advance the likelihood of an evaluation's success. Developed initially by Joseph Wholey and associates at the Urban Institute, evaluability assessment focuses first on program characteristics and then on the feasibility of conducting an evaluation study as planned. As Alkin (1975) points out, "perhaps it is important as evaluators that we learn to distinguish those situations in which the context and decision factors are so pre-determined that it can be inevitably said that no one needs it [i.e. the evaluation] and no one cares" (p. 211). At the local level, evaluability assessment can be a first step in the collaboration between evaluators and decision-makers and may result in more appropriate evaluations and, hence, better use of their results.

An evaluability assessment begins by examining characteristics of program components to determine how close they come to an ideal, asking if they are well-defined and capable of being implemented in a prescribed manner; if goals and effects are clearly specified; and if plausible causal connections link goals and effects. In this stage, the evaluator develops three models of the program: first, a "program documents model," which shows the causal links described in program materials; second, a "program manager's model," which modifies the first model according to information from key decision-makers; and, finally, an "evaluable program model," which presents the evaluator's

views of what components can appropriately be evaluated. This first step develops an explication and analysis of the theoretical premises of the program and a specification of the "process model" of the program--two methods suggested by Weiss (1972, p. 323) for increasing the use of evaluation results.

Because the purpose of the evaluation determines its methodological requirements, the second stage of the evaluability assessment, the feasibility analysis, begins by determining the purpose(s) of the given evaluation, then looks at the constraints on the evaluation to see to what extent the research requirements can be met. Considered in the feasibility analysis are program design and implementation, information requirements, and research design. Its product is a list, based on the evaluable program model, of the program components and the goals and effects to be studied in the ultimate evaluation. The program evaluation, when it is finally conducted, benefits from the limits set during the two stages of the evaluability assessment; only what can and should be evaluated will be evaluated. As Rutman summarizes,

> Constraints inevitably limit the pursuit of the study's goals. However, recognition of the practical implications of the constraints is crucial in determining the point at which the study's objective can no longer be achieved. At this point the purposes of the evaluation may have to change. Or further work may be needed to remove obstacles that inpede the measurement process, as a precondition of doing the study (p. 145).

Part of the appeal of evaluability assessment for local evaluations is due to the fact that programs themselves can benefit from the process. Before the program evaluation begins and as a by-product of the evaluability assessment, program directors may make

changes in the program to enhance its evaluability. Strategies may include analyzing problems; specifying outcomes; assessing program design and implementation; and facilitating program development (what Rutman calls "formative research"). By helping local decision-makers in this process, evaluators can educate users to the practices of good evaluation while they themselves learn the details of the functioning program.

The benefits of conducting evaluability assessments, then, fall into two categories. First, evaluablity assessments facilitate evaluation planning by establishing priorities, by providing "front-end control" over the evaluation process, and by allowing a wise allocation of evaluation dollars. Second, they facilitate planning and implementation by providing information on appropriate directions both for information users and for evaluators. To the extent that internal evaluators are "captive . . . in the sense that they do not often have a choice whether or not to do an evaluation" (Caulley and Smith, 1978, p. 31), evaluability assessments may currently, however, have limited applicability.

The preceding review of both empirical and non-empirical literature has included a variety of factors involved in local uses of evaluation information and numerous suggestions as to what evaluators should do about them. No one method can be viewed as the panacea for evaluation's utilization woes, and it is important that evaluators use a holistic approach to adopting these strategies. As Weiner, Rubin, and Sachse (1977) argue, "Attempts to increase evaluative influence which focus on a few of these factors in isolation and which do not recognize the highly complex and interactive system of forces

constraining evaluator activity are likely to fail to alter the overall effects of the system" (p. 23).

A Critique of Previous Use Research

The previous research on the use of evaluative information has been dominated by two research approaches. The first approach is that of the retrospective case study, of which a limited number have been reported (e.g. Alkin, Daillak, and White, 1979). The second type of evaluation use study involves simulation investigations that at least purport to be grounded in communications or attribution theory (e.g. Newman, Brown, and Braskamp, 1980, pp. 29-36). Simulation research typically presents administrators with a "simulated" evaluation report in which different report features, e.g. the sex of the evaluator, are varied, and the impacts of the variations then assessed. Both research approaches have made major contributions to our understanding of the use of evaluative information, but both have recognizable weaknesses.

Alkin (1979) has consistently argued that

> The forces which lead to utilization are indeed complex. This complexity in combination with our current inadequate understanding of evaluation and utilization requires a methodological procedure sufficiently sensitive to capture the nuances involved--naturalistic research is currently a most appropriate tool for a study of evaluation utilization (p. 13).

Alkin, Daillak, and White (1979) are less restrained when they argue that "the choice of appropriate research strategies can be reduced to one class: naturalistic research methods" (p. 32, emphasis in original). Although appealling and often appropriate, this case study work can be criticized on at least three grounds.

First, case study research can no longer call itself theory generating. There is now theory, or at least the beginnings of theory, that can be relied upon in conducting use research. For example, our understanding of evaluation is reflected in fairly elaborate conceptualizations of types of evaluation use (Braskamp, 1980). As noted above, several researchers have differentiated instrumental, conceptual, and symbolic uses of evaluation. Meltsner's (1976) conceptualization of "types" of evaluators also could support theoretically oriented inquiry. Finally, communcation-related theories can greatly enhance our understanding of at least the report phase of the evaluation endeavor (Brown and Newman, 1979 a). These various theoretical frames are certainly not yet fully developed, but further progress in developing theory absolutely depends upon testing and elaborating the constructs that are already at our disposal. From this point forward, failure to theoretically ground naturalistic research seems both unnecessary and unfortunate (Thompson, 1981b).

Second, most of the previous case study research has been based on post hoc interviews with evaluators and evaluation clients. Leviton and Hughes (1979) have commented on the dangers of retrospective research methods: "Given officials' faulty memories, retrospective research may be biased in favor of a few dramatic instances of use, rather than frequent but modest ones" (p. 15). Retrospective methods are economical, but their limitations must be acknowledged, and future studies should look to both the past and the present for data.

Third, some case study research can be criticized for the way it has been reported. Although some authors (Meltsner, 1976; Patton et

al., 1977) have integrated themes and case study evidence into a unified analysis, other researchers have presented case study evidence in non-integrated blocks of detail absent of any themes. Presented as unrelated case histories, such naturalistic research may represent a poor return on a necessarily extensive investment. Ironically, if an evaluator communicated evaluative data as poorly as some case study research has been communicated, the evaluative information might never be used.

The second approach to evaluation use research, the simulation study, can also be criticized on three grounds. First, although this research is typically represented as being theoretically grounded in communications or attribution theory, this grounding has up to now too frequently taken the form of name dropping the theory's title without invoking specific propositions of the cited theories. Of course, in some simulation studies tested propositions have been implicit, but good research practice would seem to dictate that even in these studies the proper primary focus of the work, i.e., theory building, ought to be explicitly discussed. To fail to discuss specific theoretical propositions is to lose some of the value of theory's existence; this is especially unfortunate since the theories do incorporate reasonably specific propositions about phenomena (see Davis and Salasin, 1975, p. 641). A sample proposition is offered by Thompson (1971) who argued that communicators should "provide rationalizations for listeners who are unwilling to admit that socially disapproved motives are responsible for their beliefs or actions" (p. 185).

A second danger in simulation research is that some

investigators will choose "samples of convenience" in their work because of the constraints of administering the simulations to field-based practitioners. It is unlikely that the business majors or education students who happen to enroll in graduate courses are representative of the individuals to whom the researchers sometimes attempt to generalize. This criticism is independent of and probably more telling than the recognition of some researchers (cf. Brown and Newman, 1979a, pp. 6-7) that simulation research may not perfectly generalize to natural ecologies.

Third, simulation research is inherently limited in that it typically focuses on the report phase of evaluation, and we know that "what happens before the final report is written will usually determine utilization" (Patton, 1978, p. 266). Still, it must be acknowledged that if we really want to increase evaluation use, then it may be necessary to emphasize all phases of the evaluation endeavor, including those that may be relatively less important in determining use.

Weiss (1979) summarizes four approaches to the study of use that differ in their starting places; i.e., one can trace the effects of evaluation by studying specific evaluation studies, by tracing the actions of specific people, by examining the various approaches to a given issue, or by following the effects of evaluations in a given organization (pp. 16-17). To date, research has centered on the first of these approaches in that both naturalistic and simulation studies typically focus on the effects of one study or one report. Expanding research to include all four types of studies should broaden the expected results.

Research Agenda for the Future

Despite some progress toward understanding use phenomena, much remains to be learned. Several priorities can be identified for future research.

1. Further conceptual work is needed to clarify the definitions of the terms use and rationality in the context of local decision-making involving evaluation information. Brown (1981) has suggested that

> We need a multi-dimensional definition of use which embraces the questions of the kinds of use, who uses the information, what information is used, and the context in which it is used, as well as the extent of usage (p. 7).

2. Brown also notes that expectations of use could be made an added dimension of the definition of use, but recommends that for now they be studied separately (p. 8). Conner (1979) has similarly suggested studying utilization goals and inputs (p. 17).

3. Studies of the process of local evaluation use are needed, and prospective case studies should be a high priority for future research. Conner (1979) makes this point when he writes:

> The absence of studies with a "current" time orientation is a serious missing link in the utilization research chain . . . This orientation is essential if we are to obtain the most accurate information about utilization. Retrospective studies, while useful, are subject to biases directly related to the type of use which has occurred (p. 16).

An assessment of the practicality of the collaborative approach to evaluation is needed. Brown (1981) suggests the need for a delineation of the steps involved in the process of use and "clarification of the decision-making (usage) context" (p. 8). Study should also be made of the effects of certain outcomes (e.g., negative

results or conflicting results) provided at certain times and at certain stages of program development. Further, it would be helpful to examine the effects of the early identification of potential users and the selection of issues of concern to them (Weiss, 1972, p. 324).

4. Additional insight into how evaluators perceive themselves and are perceived by administrators is needed if the personal factor is accepted as an important determinant of use. A broader understanding of these perceptions is needed similar to that which has been achieved in areas such as teacher education (Miller, Thompson, and Frankiewicz, 1975).

5. Local use research needs to broaden its subjects to include more principals, teachers, and school board members. Several researchers have involved persons in these roles as subjects (e.g., Kennedy, Apling, and Neumann, 1980; Thompson, 1981; Williams and Bank, 1980), but the use of people in these roles as subjects is not in proportion to the influence that they exert over program operation. Regarding principals, for example, Lipham (1980) notes that "the leadership behavior of the principal is a powerful factor which influences the adoption and institutionalization of an educational change" (p. 83).

6. Researchers need to determine whether or not school personnel can effectively be trained to make more optimal use of evaluative information. For example, research is needed to determine if it is feasible to help administrators increase "problem solving capabilities and ability to express and articulate [information] needs" (Haenn, 1980, p. 13). Training of local evaluation staff members should also be explored since "there is a striking absence of

any formal training in evaluation for staff" (Lyon et al., 1978, p. 70).

7. Given the importance of informal communication, further empirical work is needed on the effects of local dissemination efforts. As Weiss (1972) proposed, the following variables deserve examination: clarity and attractiveness of presentation; inclusion of implications for action; use of inventive mechanisms to reach remote audiences; and aggressive advocacy by evaluators (p. 325).

8. Research on the merits of mandated evaluation is also needed. Do programs that are not evaluated differ from comparable evaluated programs regarding either program processes or productivity? Also, at some point the notion of mandated evaluation itself needs to be evaluated.

9. Both Brown (1981) and Sanders (1981) have suggested broadening the theoretical basis of use research to include application of theories from other disciplines. The study of the literature on interpersonal influence, on communications theory, and on decision-making theory may well provide further insight into the use of evaluative information.

Making Local Evaluation Use Happen

People who discuss evaluation use tend inappropriately to lay the blame for non-use at someone else's doorstep. However, everyone involved in the evaluation enterprise at the local level must recognize that it takes at least two to create non-use. Administrators must assume some responsibility for making sure that evaluative information is both usable and used (Meltsner, 1976). So

too evaluators must accept some responsibility for making use happen. As Polivka and Steg (1978) argue,

> Traditionally, the evaluator has been very hesitant to claim any responsibility for the use of his/her findings. This approach has helped make it very easy to ignore evaluation results (p. 697).

This can no longer be the case. Standard A8, "Evaluation Impact" (Joint Committee, 1981), reminds us that "evaluators must not assume that improvements will occur automatically once the evaluation report is completed" (p. 47). Both administrators and evaluators need to recognize that the responsibility for use is not a "zero-sum" game in which responsibility can be divided up and will always total 100%. They would be better off if both administrators and evaluators assumed 60 or 80% of the responsibility for increasing use. Even then, as Patton (1978) notes, "Increasing utilization potential does not guarantee utilization of findings. There are no guarantees" (p. 96).

Bibliography

Alkin, M.C. Evaluation: Who needs it? Who cares? Studies in educational evaluation, 1975, 1, 201-212.

Alkin, M.C. The title of this speech is . . . Paper presented at the Annual Meeting of the California Educational Research Association, Burlingame, CA, 1976. (ERIC No. 160 652)

Alkin, M.C. Using naturalistic research for the study of evaluation utilization. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1979. (ERIC No. 170 378).

Alkin, M.C. Improving Title I technical assistance: An evaluation utilization perspective. Paper presented at the Annual Meeting of the American Educational Research Association, 1980. (a)

Alkin, M.C. Naturalistic study of evaluation utilization. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass, 1980. (b)

Alkin, M.C. & Daillak, R.H. A study of evaluation utilization. Educational evaluation and policy analysis, 1979, 1, 41-49.

Alkin, M.C., Daillak, R.H., & White, P. Using evaluations: Does evaluation make a difference? Beverly Hills: Sage Publications, 1979.

Alkin, M.C., Kosecoff, J.B. A new eclectic model for the redirection of evaluation efforts. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, 1973.

Alkin, M.C., Kosecoff, J.B., Fitz-Gibbon, C., & Seligman, R. Evaluation and decision-making: The Title VII experience. CSE monograph series in evaluation, 1974, No. 4.

Andrews, J.V. Reactions to program evaluation: A qualitative analysis (Doctoral dissertation, Cornell University, 1979). Dissertation abstracts international, 1979, 39, 6717A. (University Microfilms No. 79-10, 731).

Bank, A., Snidman, N., & Pitts, M. Evaluation, dissemination, and educational improvement: How do they interact? Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1979.

Bosco, J. The role of the administrator in the improvement of evaluation studies. Education, 1971, 92, 70-74.

Brandl, J.E. Policy evaluation and the work of legislatures. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Braskamp, L.A. Assessing the utilization of a program evaluation: A review. Paper presented at the conference, Planning and Conducting Program Evaluations and Reviews in Higher Education, St. Petersburg, 1980.

Braskamp, L.A. & Brown, R.D. (Eds.) Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Braskamp, L.A., Brown, R.D., & Newman, D.L. Credibility of a local educational program evaluation report: Author source and client characteristics. American educational research journal, 1978, 15, 441-450.

Brickell, H.M., Aslanian, C.B., & Spak, L.J. Data for decisions: An analysis of evaluation data needed by decision-makers in educational programs. New York: Policy Studies in Education, 1974.

Brown, E.K. Changes in school management's needs for evaluation information. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, 1973. (ERIC No. ED 074 094).

Brown, R.D. Evaluation utilization: A literature review and research agenda, A critique. Unpublished manuscript, 1981.

Brown, R.D., & Braskamp, L.A. Summary: Common themes and a checklist. In L.A. Braskamp & R.D. Brown (Eds.) Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Brown, R.D., & Newman, D.L. Communication theory as a paradigm for making evaluation reports more effective. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1979. (a)

Brown, R.D., & Newman, D.L. A schematic approach to studying evaluation utilization. Paper presented at the Annual Meeting of the Evaluation Research Society, Minneapolis, 1979. (b)

Brown, R.D., & Newman, D.L. An investigation of the effect of different data presentation formats and order of arguments in a simulated adversary evaluation. Education evaluation and policy analysis, in press.

Brown, R.D., Newman, D.L., & Rivers, L. Perceived need for evaluation as an influence on evaluator's impact on

decision-making. Educational evaluation and policy analysis, 1980, 2, 67-73.

Brzezinski, E.J. Quality in reporting evaluation results: A field test of the Joint Committee's Draft Standards for Educational Evaluation. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1979.

Caplan, N. What do we know about knowledge utilization? In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Caplan, N., Morrison, A., & Stambaugh, R.J. The use of social science knowledge in policy decisions at the national level: A report to respondents. Ann Arbor: Center for Research on Utilization of Scientific Knowledge, Institute for Social Research, University of Michigan, 1975. (ERIC No. 111 729)

Caulley, D., & Smith, N.L. Field Assessment Survey, No. 10. Portland, OR: Northwest Regional Educational Laboratory, 1978.

Cohen, D.K. Politics and research: Evaluation of social action programs in education. In C.H. Weiss (Ed.), Evaluating social action programs: Readings in social action and education. Boston: Allyn & Bacon, 1972.

Conner, R.F. Measuring the utilization of evaluation findings: Description and critique of different techniques. Paper presented at the Annual Meeting of the Evaluation Research Society, Minneapolis, 1979.

Cook, T.D. Utilization, knowledge-building, and institutionalization: Three criteria by which evaluation research can be evaluated. In T.D. Cook et al. (Eds.), Evaluation studies review annual (Vol. 3). Beverly Hills: Sage Publications, 1978.

Cook, T.D., & Pollard, W.E. Guidelines: How to recognize and avoid some common problems of misutilization of evaluation research findings. Evaluation, 1977, 4, 161-164.

Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F., & Weiner, S.S. Toward reform of program evaluation. San Francisco: Jossey-Bass, 1980.

Datta, L. O thou that bringest the tidings to lions: Reporting the findings of educational evaluations. Paper presented at the Annual Johns Hopkins University National Symposium on Educational Research, Baltimore, 1979.

David, J. Local uses of Title I evaluations. Educational evaluation and policy analysis, 1981, 3, 27-39.

Davis, H.R., & Salasin, S.E. The utilization of evaluation. In E.L. Struening & M. Guttentag (Eds.), Handbook of evaluation research (Vol. 1). Beverly Hills: Sage, 1975.

Deal, T.E., & Rallis, S.F. Promoting interaction among producers and users of educational knowledge. In Interorganizational arrangements for collaborative efforts. Portland, OR: Northwest Regional Laboratory, 1980.

Denny, T. How might evaluation reports be? Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980.

Dickey, B. Effective problem-solving for evaluation utilization. Paper presented at the Annual Meeting of the Evaluation Research Society, Minneapolis, 1979. (a)

Dickey, B. Utilization of evaluations of small scale educational projects (Doctoral dissertation, University of Minnesota, 1979). Dissertation abstracts international, 1979, 40, 3245A. (University Microfilms No. 79-26, 115). (b)

Dickey, B. Utilization of evaluations of small scale educational projects. Educational evaluation and policy analysis, 1980, 2, 65-77.

Eisner, E.W. Forms of representation in educational evaluation. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980.

Fullan, M. Conceptualizing problems of curriculum implementation. Paper presented for the Symposium on Curriculum Inquiry in Canada, Victoria, 1979.

Fullan, M. An R&D prospectus for educational reform. In Interorganizational arrangements for collaborative efforts. Portland, OR: Northwest Regional Educational Laboratory, 1980.

Goldberg, M. New York City program administrators' perceptions of the helpfulness of outside mandated evaluations on administrative decision-making in ESEA programs. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, 1978. (ERIC No. 156 720).

Gorham, W. Getting into action. Policy sciences, 1970, 1, 175.

Granville, A.C. An experimental study of factors that influence school decisions: Implications for evaluation. Paper

presented at the Annual Meeting of the American Educational Research Assocation, New York, 1977. (ERIC No. 139 824)

Granville, A.C. The impact of the project developmental continuity evaluation. In A.C. Granville, J.M. Love, R.D. Matz, L.J. Schweinhart, & A.G. Smith (Eds.), The impact of evaluation: Lessons drawn from the evaluations of five early childhood education programs. Proceedings of a symposium presented at the Annual Meeting of the American Educational Research Association, Toronto, 1978. (ERIC No. 166 212)

Gray, P.J. The development and testing of a collaborative approach to education program evaluation (Doctoral dissertation, University of Oregon, 1979). Dissertation Abstracts International, 1980, 40, 4966A. (University Microfilms No. 80-5, 767).

Grobe, R.P. Evaluation--What's it all about? NASSP Bulletin, 1978, 62, 1-14.

Guba, E.G. An overview of the evaluation problem. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, 1969. (ERIC No. 031 814)

Guba, E.G., & Stufflebeam, D.L. Evaluation: The process of stimulating, aiding, and abetting insightful action. Bloomington: Indiana University, 1970. (ERIC No. 055 733)

Guskin, A.E. Knowledge utilization and power in university decision making. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Guttentag, M., & Struening, E.L. The handbook: Its purpose and organization. In M. Guttentag & E.L. Struening (Eds.), Handbook of evaluation research (Vol. 2). Beverly Hills: Sage, 1975.

Haenn, J.F. Reasons why evaluations and testing don't inform. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980. (ERIC No. 187 733)

Hall, G., & Loucks, S. A developmental model for determining whether the treatment is actually implemented. American educational research journal, 1977, 14, 263-276.

Haller, E. Cost analysis for educational program evaluation. In W.J. Popham (Ed.), Evaluation in education. Berkeley: McCutchan, 1974.

Hamilton, G.E. Title I evaluation from the perspective of

the local school. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980.

Hansen, J.B., & Martin, J.M. Annotated bibliography on the use of evaluation. Portland, OR: Northwest Regional Educational Laboratory, 1980.

Havelock, R.G. Dissemination and translation roles. In T.L. Eidell & J.M. Kitchel (Eds.), Knowledge production and utilization in educational administration (Columbus, Ohio) and Center for the Advanced Study of Educational Administration (Eugene, Oregon) Career Development Seminar, October, 1967. Eugene, OR: University of Oregon Press, 1968.

Hayman, J. The case for internal linkage: Functional integration within the organization. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1979.

Holley, F.M. Catch a falling star: Promoting the utilization of research and evaluation findings. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1979.

Holley, F.M. Evaluation utilization: Is it easier to move a mountain than a molehill? Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980. (a)

Holley, F.M. What it takes to win: Factors in the utilization of evaluation findings for educational improvement. In C.B. Stalford (Ed.), Testing and evaluation in schools: Practitioners' views. Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, and the National Institute of Education, October, 1980, 101-111. (b)

Johnson, M.D. Evaluation as a model for decision-oriented research. Paper presented at the Association for Institutional Research Annual Forum, 1978. (ERIC No. 154 743)

Johnston, J. What we know about the researcher and research methods in evaluation. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, 1978.

Joint Committee on Standards for Educational Evaluation. Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill, 1981.

Kelezo, G.J. Program evaluation: Emerging issues of possible legislative concern relating to the conduct and use of evaluation in the Congress and executive branch. Washington, D.C.: Congressional Research Service, Library of Congress, 1974.

Kennedy, M.M., Apling, R., & Neumann, W.F. The role of evaluation and test information in public schools. Cambridge, MA: The Huron Institute, 1980.

Kilbourne, R., & DeGracie, J. The use of L.E.A. research at the local level: The picture of a dropout. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1979.

King, J.A., & Thompson, B. A nationwide survey of administrators' perceptions of evaluation. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, 1981.

King, J.A., Thompson, B., & Pechman, E.M. Evaluation utilization: A bibliography. New Orleans: Orleans Parish Public Schools, 1981. (a)

King, J.A., Thompson, B., & Pechman, E.M. Evaluation utilization: An annotated bibliography. New Orleans: Orleans Parish Public Schools, 1981. (b)

Knorr, K.D. Policymakers' use of social science knowledge: Symbolic or instrumental? In C.H. Weiss (Ed.), Using social research in public policymaking. Lexington, MA: Heath, 1977.

Law, A. In M.C. Alkin & A. Law, A conversation on evaluation utilization. Educational evaluation and policy analysis, 1980, 2, 73-79.

Lee, A.M., & Holley, F.M. Communicating evaluation information: Some practical tips that work. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, 1978.

Leviton, L.C., & Hughes, E.F.X. Utilization of evaluations: A review and synthesis. Evanston, IL: Center for Health Services and Policy Research, Northwestern University, 1979.

Lipham, J.M. Change agentry and school improvement: the principal's role. In Interorganizational arrangements for collaborative efforts. Portland, OR: Northwest Regional Educational Laboratory, 1980.

Lyon, C.D., Doscher, L., McGranahan, P., & Williams, R. Evaluation and school districts. Los Angeles: Center for the Study of Evaluation, 1978.

Meltsner, A.J. Policy analysts in the bureaucracy. Berkeley: University of California Press, 1976.

Miller, A.H., Thompson, B., & Frankiewicz, R.G. Attitudes of teacher education students toward teachers. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., 1975. (ERIC No. 104 887)

Mitchell, D.E. Social science impact on legislative decision-making: Process and substance. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980. (ERIC No. 186 347)

Newman, D.L., & Brown, R.D. An exploratory investigation of decision-making style and locus of control as factors related to choice of school evaluation strategies. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980.

Newman, D.L., Brown, R.D., & Braskamp, L.A. Communication theory and the utilization of evaluation. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Patton, M.Q. Utilization-focused evaluation. Beverly Hills: Sage, 1978.

Patton, M.Q., Grimes, P.S., Guthrie, K.M., Brennan, N.J., Grench, B.D., & Blyth, D.A. In search of impact: An analysis of utilization of federal health evaluation research. In C.H. Weiss (Ed.), Using social research in public policymaking. Lexington, MA: Heath, 1977.

Pelz, D.C. Some expanded perspectives on use of social science in public policy. In J.M. Yinger & S.J. Cutler (Eds.), Major social issues, a multidisciplinary view. New York: The Free Press, 1978.

Polivka, L., & Steg, E. Program evaluation and policy development: Bridging the gap. Evaluation quarterly, 1978, 2, 696-707.

Randall, R.S. Knowledge about decision processes and information. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, 1969. (ERIC No. 031 814)

Revicki, D., & Rubin, R. Models for measuring program implementation: A review and critique. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980.

Rich, R.F. Uses of social science information by federal bureaucrats: Knowledge for action versus knowledge for understanding. In C.H. Weiss (Ed.), Using social research in

public policymaking. Lexington, MA: Heath, 1977.

Rippey, R.M. Studies in transactional evaluation. Berkeley: McCutchan, 1973.

Ross, J.A. Decision rules in program evaluation. Evaluation review, 1980, 4, 59-74.

Rossi, P.H. Booby traps and pitfalls in the evaluation of social action programs. In C.H. Weiss (Ed.), Evaluating action programs: Readings in social action and education. Boston: Allyn and Bacon, 1972.

Rutman, L. Planning useful evaluations. Beverly Hills: Sage, 1980.

Saks, M.J. The utilization of evaluation research in litigation. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Sanders, J. A critique of "Evaluation utilization: A literature review and research agenda" by Bruce Thompson and Jean A. King. Unpublished manuscript, 1981.

Scriven, M. The methodology of evaluation. In B.R. Worthen, & J.R. Sanders (Eds.), Educational evaluation: Theory and practice. Worthington, OH: Charles A. Jones Publishing Company, 1973.

Shapiro, E. Educational evaluation: Rethinking the criteria of competence. School review, 1973, 81, 523-549.

Shapiro, E. Assessing the impact of evaluation: A univariate time series technique. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1979. (ERIC No. 174 689)

Simon, H.A. Administrative behavior. New York: MacMillan, 1957.

Smith, N.L. Bibliography on evaluation utilization, No. 39. Portland, OR: Northwest Regional Educational Laboratory, 1980. (a)

Smith, N.L. Evaluation utility and client involvement in accreditation studies. Educational evaluation and policy analysis, 1980, 2 (5). (b)

Smith, N.L. The president's corner. Evaluation news, Summer, 1980, 24-25. (c)

Stake, R. Evaluation design, instrumentation, data

collection, and analysis of data. In B.R. Worthen & J.R. Sanders (Eds.), Educational evaluation: Theory and practice. Worthington, OH: Charles A. Jones, 1973.

Stake, R. Evaluation design, instrumentation, data collection and analysis of data. Urbana, IL: Center for Instructional Research and Curriculum Evaluation, 1976.

Stevenson, J.F. Evaluation utilization in human service organizations: Is there a role for the in-house evaluator in assisting utilization? Paper presented at the Annual Meeting of the Evaluation Research Society, Minneapolis, 1979.

Stufflebeam, D.L., & Webster, W.J. An analysis of alternative approaches to evaluation. Educational evaluation and policy analysis, 1980, 2, 5-20.

Thompson, B. Administrators' perceptions of various evaluation report styles. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, 1981. (a)

Thompson, B. Review of Utilization of evaluative information by L.A. Braskamp & R.D. Brown (Eds.) Educational evaluation and policy analysis, 1981, 3, 106-107. (b)

Thompson, B. Validity of an evaluator typology. Educational evaluation and policy analysis, 1980, 2, 59-65.

Thompson, B., & King, J.A. Evaluator types: Do differences matter? Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, 1981.

Thompson, W.N. Modern argumentation and debate. New York: Harper and Row, 1971.

Webster, W., & Stufflebeam, D. The state of theory and practice in educational evaluation in large urban school districts. Invited address presented at the Annual Meeting of the American Educational Research Association, Toronto, 1978.

Weiner, S.S., Rubin, D., & Sachse, T. Pathology in institutional structures for evaluation and a possible cure. Stanford, CA: Stanford Evaluation Consortium, 1977.

Weiss, C.H. Utilization of evaluation: Toward comparative study. In C.H. Weiss (Ed.), Evaluating action programs: Readings in social action and education. Boston: Allyn & Bacon, 1972.

Weiss, C.H. (Ed.). Using social research in public policy making. Lexington, MA: Lexington Books, 1977.

Weiss, C.H. Conceptual issues in measuring the utilization of research and evaluation. Paper presented at the Annual Meeting of the Evaluation Research Society, Minneapolis, 1979.

Wholey, J.S., Scanlon, J., Duffy, H., Fukumoto, J., & Vogt, L. Federal evaluation policy: Analyzing the effects of public programs. Washington, D.C.: The Urban Institute, 1970.

Wilensky, H.L., & Lebeaux, C.N. Industrial society and social welfare. New York: Russell Sage, 1958.

Williams, R.C., & Bank, A. School district linking of evaluation and testing with instruction: A problem worthy of attention. Paper presented at the Annual Meeting of the Evaluation Research Society, Arlington, Virginia, 1980.

Wise, R.I. What we know about the decision-maker and decision settings. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, 1978.

Wise, R.I. The evaluator as educator. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Worthen, B.R., & Sanders, J.R. Educational evaluation: Theory and practice. Worthington, OH: Charles A. Jones, 1973.

Young, C.J., & Comptois, J. Increasing Congressional utilization of evaluation. In F. Zweig (Ed.), Evaluation in legislation. Beverly Hills: Sage, 1979.

Zepeda, R.A. What the models cannot tell us: Information needs for program planning. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980.